Is Marriage A Normal Good?
Evidence from NBA drafts

Jiaqi Zou*

December 7, 2021

Click here for the most recent version

Abstract

Despite globally declining marriage rates in recent decades, little is known about whether improvements in male economic status increase marriage. This paper tackles lack of data on *permanent* income shocks for men by examining a natural experiment surrounding the NBA's annual player drafts. I exploit two institutional features: well-defined initial salaries decreasing monotonically by draft order and high-quality draft predictions that inform player expectations. To isolate the causal effect of male earnings on marriage outcomes, I show that disparities between predicted and actual draft ranks exogenously shift player salaries. This setup provides novel income treatments that are not only large and individual-specific but also opportunely occurring early in career and adult life, before family formation takes place. Constructing a new dataset tracking players' major family decisions, I am the first to show men are indeed more likely to marry when their earnings increase, despite modern-day normalization of cohabitation. For the 2004-2013 draft cohorts, a 10% increase in initial five-year salary raises likelihood of marriage by 8.9%. Excluding superstar draft picks yields larger and more significant results, reasonably suggesting lower income men are *more* responsive to income shocks.

*I am grateful to my advisors, Loren Brandt, Kory Kroft, Laura Turner, and Román Zárate for their guidance and encouragement on this project. For very helpful comments, I would also like to thank Emily Breza, Chaoran Chen, Victor Dee, Kevin Devereux, Jiaying Gu, Jonathan Hall, Kevin Lim, Yao Luo, Étienne Makdissi, Ismael Mourifié, Serdar Ozkan, Jeffrey Penney, David Price, Swapnika Rachapalli, Baxter Robinson, Heather Sarsons, Marc-Antoine Schmidt, Hannes Schwandt, Hammad Shaikh, Aloysius Siow, Michael Stepner, Clémentine Van Effenterre, and Hanzhe Zhang, as well as numerous seminar participants at the University of Toronto's summer workshop and applied micro lunch seminar. Contact: Department of Economics, University of Toronto; email: jiaqi.zou@mail.utoronto.ca.

1 Introduction

With marriage rates declining globally in recent decades (Schoen and Canudas-Romo, 2005; Schoen and Cheng, 2006; Kalmijn, 2007),[1] survey evidence indicates over two-thirds of never-married U.S. adults cite financial reasons for being as of yet unmarried (Pew Research Center, 2017, 2019).[2,3] Despite the hypothetical self-reported nature of this result, the idea that people *do* want to marry but may be constrained from following through—either by financial barriers or not having attained a self-imposed threshold of financial success—calls for further investigation. That is, how would a change in income *actually* affect marriage decisions? While existing literature has established that women who receive positive income shocks tend to *delay* marriage (Baird et al., 2011; Teitler et al., 2009; Hankins and Hoekstra, 2011),[4] empirical evidence on the causal link between male economic status and marriage outcomes has proved elusive, owing to scant data on unanticipated *permanent* income shocks for men (Hankins and Hoekstra, 2011; Kearney and Wilson, 2018).

This paper tests the classic question of whether marriage is a normal good, targeting the literature gap for men. To examine whether changes in men's earnings affect their propensity to marry, I exploit NBA drafts as a natural experiment, where unanticipated changes in player draft order generate exogenous income shocks. A key novelty of this setup is it affords a rare opportunity to work with income treatments that are not only large and individual-specific but also opportunely occurring early on in both career and adult life, before marriage decisions take place. I am the first to show causal evidence that men are indeed more likely to marry when their earnings increase, despite modern-day normalization of cohabitation. For the 2004-2013 draft cohorts, I find that a 10% increase in initial five-year salary raises likelihood of marriage by 8.9%, both on average and consistently for positive and negative income shocks. Restricting away from superstar draft picks yields larger and more significant results which, consistent with reasonable priors, suggests that lower income men are *more* responsive to income shocks.

[1] Individuals are marrying both later and at lower rates overall. U.S. median age at first marriage rose steadily from just above age 20 in the late 1960s to around age 30 today, and ever-married rates fell for all age categories up to 64 since 1970 (Lee and Payne, 2010; U.S. Census Bureau, 2020).

[2] This statistic is based on the 85% of never-married adults who indicated they wanted to marry someday or were undecided and excludes the 15% who indicated they did not want to marry.

[3] Financial reasons differ by individual, from wanting the ability to provide for a future family to prioritizing career attainment, but apply to men and women alike.

[4] For our purposes, I focus exclusively on the literature examining *unconditional* cash transfers. This excludes conditional cash transfer papers, such as Field et al. (2016) and Salisbury (2017), as receipt of treatment is *conditioned* on marriage outcome, which combines underlying preferences for marriage with preferences for money.

Four features of the NBA data allow me to test whether unanticipated changes in salary affect propensity to marry. First, prospective players are drafted into the NBA in order from 1 to 60. Crucially, initial contract sizes are well-defined and decrease monotonically by draft order, affecting the first five years of salaries. This provides a direct one-to-one mapping of draft ranks into salary expectations. Second, I exploit a pre-draft tradition, where well-respected experts publish high-quality predictions about the order in which the prospective players will most likely be drafted. This serves as an individual-specific estimate of each player's initial earning potential, informing player expectations leading up to draft night. Third, draft night is subject to substantial randomness, owing to unanticipated team tie-breaking between players. This provides unexpected changes in player draft order, as measured by the disparity between each player's predicted and actual draft rank. Finally, a combination of the basketball players' public figure status and the U.S. Freedom of Information Act allow me to track their major family decisions over time. I accomplish this using a variety of primary sources, such as interviews and social media, as well as crowdsourced police reports and court documents.

To study the causal effect of salary on propensity to marry, I use prospective players' rank shock as an instrument for salary. Without an IV, a naïve regression of marital status over salary yields a coefficient that is not statistically significant, with or without controls. As salary is endogenous, however, the true causal effect could still be positive or negative, depending on the net effect of omitted variable biases and reverse causality. To use rank shock as an IV, I argue that it satisfies three assumptions: a strong relationship with salary, random assignment on both potential earnings and marriageability characteristics, and the exclusion restriction. First, I obtain a strong first stage with an F-stat well above 10 and a robust relationship between rank shock and salary that remains unchanged, with or without the addition of controls. Second, rank shock is randomly assigned across players in theory, as it stems from the randomness of teams tie-breaking unexpectedly between players. Reduced-form checks of rank shock's relationship with *observable* marriageability characteristics and earnings potential yield no statistical significance across the board. Third, I argue that rank shock only affects players' propensity to marry by changing the size of their initial contract. Although rank shock technically also changes the player's NBA team and therefore the location to which he is assigned, internalizing this in the first stage does not meaningfully change the causal effect of interest. Satisfying these assumptions, I can employ two-stage least squares to obtain average causal effects.

I am the first to show causal evidence that men who experience positive earnings shocks are indeed more likely to marry, in spite of the gradual normalization of cohabitation in modern-day society. I am also the first to ascertain that the causal relationship is consistent across positive and negative income shocks within the same sample. A main difficulty in the literature is finding a natural experiment that can provide both sufficiently large income shocks and clean identification for men. While there are plenty of female-targeted interventions, such as cash incentives to reduce child marriage in developing nations (as in Baird et al. (2011)), and government income subsidies for single mothers in developed nations (as in Teitler et al. (2009)), there is no male-targeted policy equivalent. Men may also require larger pecuniary shocks than women before an effect can be observed on their marriage outcomes. Comparing lottery winners of $25,000 versus $50,000 cash prizes, for example, Hankins and Hoekstra (2011) finds that, while women who won the larger prize tended to marry later, there was no differential effect between the male winners. Larger shocks tend to be economy-wide shocks, which impact men and women simultaneously, posing a challenge to clean identification. A notable exception is Autor et al. (2019), which employs (1990-2010) import shocks to U.S. male manufacturing employment to show that marriage rates fall in response to negative male earnings shocks. Kearney and Wilson (2018) sets out to show the corresponding result for positive income shocks using the (1997-2012) U.S. fracking boom. However, the authors find no evidence of higher marriage rates in areas with greater geographical exposure to shale plays and fracking production. While they further replicate their methodology using the (1969-1987) Appalachian coal boom and do find an increase in marriages in this earlier setting, they hypothesize that marriage may *no longer* be a normal good on account of recent social norm shifts regarding cohabitation. I avoid many of the data challenges in the literature with a natural experiment design surrounding the (2004-2013) NBA drafts, which provides rich variation in individual-specific income treatments that occur before family formation decisions are made. I fill in the literature gap for positive income shocks and confirm Autor et al. (2019)'s result for negative income shocks. Thus, I am able to tell a unifying story that marriage is indeed a normal good for my sample, consistently across the domain of positive and negative income shocks.

Further evidence suggests my result constitutes a *lower-bound* estimate for general population men. First, I show my finding is not driven by the superstars. Dropping the top draft picks, where income differentials and any differences in ability and fame are the largest, I find the result persists for the restricted, more homogeneous subsample. If anything, effect sizes are larger and more significant for those of lower expected salary. That we should expect the magnitude to increase and not decrease is supported by both empirical findings and intuition. For comparability with what we know about the general population, I convert my results into income elasticities of marriage demand, η.

For the 2004-2013 cohorts, my result translates to η_{NBA}=0.18 for the overall sample[5] and increases to 0.65 for the subsample drafted outside of the top 45. This fits in nicely with Kearney and Wilson (2018)'s (1969-1987) Appalachian coal boom $\eta_{Appalachia} = 0.96$ before shifts in social norms regarding cohabitation took place. While marriage is a normal good in all three cases, it borders on being a luxury good for general population men who, as expected, are most responsive to changes in income. Further intuition to help rule out marriage as an inferior good is that more money is more power to live life according to one's true preferences. If the underlying preference were bachelorhood, then greater financial freedom should better facilitate this lifestyle, such as with purchasing prepared meals, hiring housekeeping, etc. If this were the case, we should see a significant negative relationship between salary and marriage, but the opposite is true. Finally, the main reasons for why people marry can largely be classified into three broad categories: love and companionship, social conformity pressures, and financial and practical considerations (Pew Research Center, 2019), in order of response popularity. I argue the latter two if not all three channels are weaker for basketball players. Not only that, but NBA players also face substantially higher costs in the event of a divorce. Taken together, basketball players should have *lower* incentives to marry than general population men. This reaffirms that, if marriage remains a normal good for my sample *in spite of* these factors, we should only expect to see a greater response for general population men. Additionally, I extend my external validity results to other major family decisions, such as propensity to have children and number of children, and confirm that children are normal goods, as in Black et al. (2013).

I then investigate whether the income-induced changes in marriage rates are driven by the man's demand or his partner's. Suggestive evidence indicates it is the man who is incentivized to marry by his earnings shock. To see this, I test a tabloid theory that NBA players are vulnerable to "baby trapping" – i.e., the partner may attempt to increase their chances of marrying the NBA player by first having a baby with him. I find, however, that players who receive positive income shocks can effectively use this additional bargaining power to *defer* babies until after he is ready to commit in terms of engagement or marriage. Not only that, but these income-induced marriages also tend to be high-quality unions not associated with meaningful changes in divorce. Taken together, this constitutes encouraging evidence that, contrary to media speculation, basketball players are not being persuaded into marriage through children but choosing to marry in their own time and on their own terms.

Finally, I employ the seminal Becker (1973, 1974)-Keeley (1977) framework to tie my

[5]For the overall sample, a 100% increase in initial *five-year* earnings increased marriage rates by 89%. Thus, a 100% increase in *annual* earnings increased marriage rates by 17.8%, yielding an elasticity of 0.178.

results in with those of the literature. Namely, their model of household specialization predicts that changes in earnings that improve the economic gains to marriage induce greater propensity to marry and earlier marriage. Conversely, changes in earnings that erode the gains to marriage reduce likelihood of marriage and delay marriage. That we observe female recipients of positive income shocks tending to delay marriage yet male recipients of positive income shocks tending to embrace it is consistent with the theory, given the gender wage gap.

The remainder of this paper proceeds as follows. Section 2 overviews the empirical setting of NBA drafts. Section 3 describes the four datasets I collect and compile. Section 4 outlines the empirical framework for causal inference. Sections 5 and 6 present the results of the two-stage least squares and alternative specification, respectively. Section 7 contains robustness checks. Section 8 ties the results together with the classic Becker-Keeley framework. Section 9 concludes.

2 Empirical Setting

I employ a natural experiment surrounding NBA drafts to overcome lack of data on unanticipated permanent income shocks for men. This section gives a brief overview of the institutional setting and the four features of the NBA that this setup relies on: a well-defined monotonically decreasing salary structure by draft order, high-quality individual-specific draft projections, draft night randomness, and family data availability of players over time.

2.1 Institutional Background

The **NBA draft** is a selection event held annually at the end of June, where teams in the National Basketball Association (NBA) take turns[6] drafting 60 young basketball players who are eligible and wish to join the league as professional NBA players. Prospective players are chosen in order from 1 to 60, where the first draft pick is the most coveted spot (e.g., LeBron James, 2003). Typically, 75% of prospective players are top American college basketball players and 25% are top international basketball players, and the players are usually between 20 to 24 years of age.

For salaries, the NBA publicly discloses a pre-determined rookie scale[7], which is a

[6] The order in which NBA teams select players is determined as follows: A lottery determines the first three teams to pick from among the teams that missed the previous year's finals. The remainder of the teams pick in reverse order of their record in the previous season.

[7] See, for example, the 2020-2021 rookie scale.

soft guideline[8] for how much an NBA team can pay each draft pick for the first five years of his NBA career. While salaries are explicitly stated for the top 30 draft picks to be decreasing monotonically by draft order, the monotonicity extends to the remainder of the draft picks as well as undrafted players. This can be seen in Figure 1, which plots the five-year salaries of individual players from draft cohorts 2003-2013, denominated in 2020 USD for comparability across cohorts. The fitted curve decreases smoothly across all 60 draft picks as well as undrafted hopefuls, whom I denote as draft pick 61. Since players' coaches and agents know the expected salaries at each draft rank from industry experience and provide their player with the relevant information, the log-quartic curve of best fit[9] estimates those values. This fitted curve thus provides a direct one-to-one mapping of draft ranks into salary expectations.

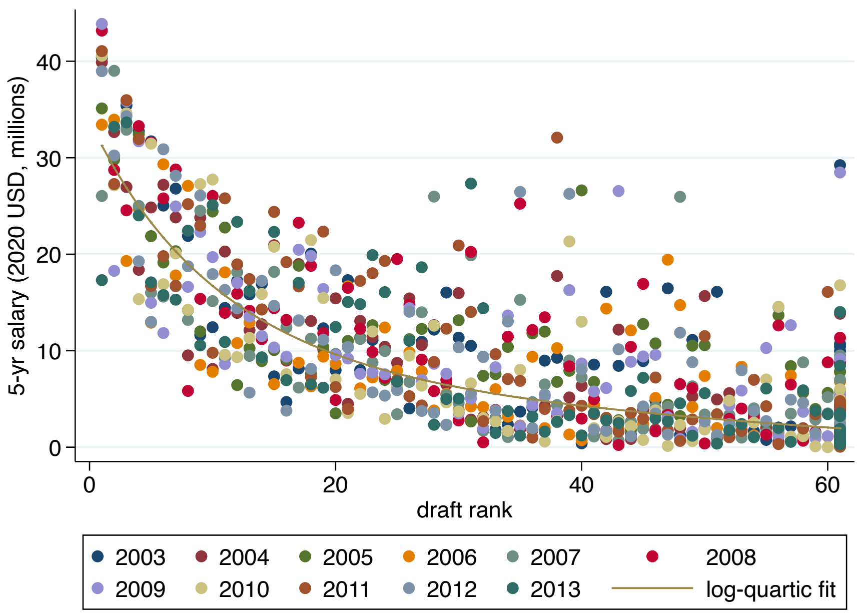


Figure 1: **Five-year salaries by draft pick**. Each dot represents the realized five-year salary of a basketball player, denominated in 2020 USD, plotted by draft rank, and color-coded to reflect draft year. The log-quartic curve of best fit estimates the expected salary for each draft pick based on 11 cohorts of individual salary data.

I exploit a pre-draft tradition, where a month prior to draft night, mainstream sports writers and analysts publish their prediction of player draft order x_i^e, known as their

[8]The NBA teams have discretion to offer as little as 80% to as much as 120% of the scale amount.

[9]The model is expressed as $\hat{y} = exp(\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3 + \hat{\beta}_4 x^4)$, where fitted values $\hat{y}$ estimate expected salary given overall draft pick, $E[y|x]$, and x represents overall pick. This is fitted to data by regressing $ln(y)$ over a quartic function of x, collecting the estimated parameters to predict $ln(\hat{y})$, and solving for $\hat{y} = exp(ln(\hat{y}))$.

mock draft. Crucially, with their reputations on the line, their objective is to minimize forecasting error. Thus, mock drafts are *not* the order in which the experts txhink the players should be selected but their expert opinion of the *likeliest* ordering of players to be realized on draft night. While anyone could technically publish their own mock draft, an important trade-off here is having a balance of opinions while ensuring that the mocks contain top-quality information. I therefore use the projections of two well-known mock drafters, Chad Ford (of ESPN) and Aran Smith (of NBAdraft.net), who are generally recognized as expert mockers over the years relevant for my analysis.[10] Their combined judgment provides players with an individual-specific estimate of their draft placement, and thereby their initial earning potential, informing player expectations leading up to draft night.

Draft night is associated with substantial randomness. While NBA teams enter draft night with an idea of which player they want to select at their turn to draft, nothing is guaranteed because, especially outside of the top three draft picks, the teams themselves cannot perfectly anticipate the other teams' choices. When earlier drafting teams tie-break unexpectedly between players,[11] this sends later drafting teams scrambling, because the player they originally had their eye on may suddenly be taken or an even better player may suddenly become available.[12] This causes a cascading effect as later draft teams deviate from their own anticipated strategy. Upon being picked, each player realizes his actual draft rank x_i^a. Given the NBA's well-defined initial salary structure, x_i^e and x_i^a are associated with different initial contracts and five-year salary trajectories. In theory, this provides random assignment of rank shocks to players, which exogenously shift player salary.

Finally, by virtue of the players' public figure status, family updates such as engagements, weddings, and baby births are periodically provided through a variety of public sources. These include media interviews, congratulatory announcements by their team, or celebratory announcements on their own social media. Occasionally, police reports and court

[10] See, for example, Business Insider (2011)'s evaluation of the mock drafts published by representatives of eight mainstream sports sites/newspapers, using a scorecard constructed on average error per pick, number of correct picks, and number of predicted top 30 players taken, of which Chad Ford and Aran Smith ranked first and second, respectively. Anecdotally, they are also referenced in interviews with basketball players.

[11] Unexpected tie-breaks occur even within the top three draft picks. For example, the Golden State Warriors reportedly broke their promise to draft LaMelo Ball with their 2$^{\text{nd}}$ pick and ended up choosing James Wiseman.

[12] While teams may signal interest to players and their agents beforehand, it is well-known that anything can happen on draft night. For example, the Chicago Bulls notified Gorgui Dieng of their interest in selecting him with their 20$^{\text{th}}$ pick. However, on the clock, the team's front office decided to choose Tony Snell instead, who was ranked higher on their preference list. Additionally, Darington Hobson explains on Kennedy, Alex (host) (2017) that, while six teams picking in the first round showed interest, he was ultimately drafted 39$^{\text{th}}$.

documents supplement divorce and child birth dates. The partner's details can generally be found in player fan pages or public records searches, after which it can be corroborated or expanded upon from the data provided on the partner's own social media.

3 Data

I collect and compile a novel dataset that contains publicly available data on expert draft projections, draft outcomes, annual salaries, and family information for the universe of NBA draft picks and mocked-but-undrafted players for the 2004-2013 draft cohorts.

3.1 NBA drafts: actual and mock rankings

All historical NBA draft outcomes can be found on NBA.com. Each cohort year is comprised of 60 drafted players, with the exception of the 2004 draft, which has 59 players. Players' draft rank is determined by the order in which they were selected on draft night. Additional information includes player affiliation[13] and draft team.

For draft projections, I employ the finalized mocks of NBAdraft.net (by Aran Smith) and ESPN (by Chad Ford). ESPN and NBAdraft.net projections exhibit a correlation of 0.843 and 0.862 with actual draft rank, respectively. Taken together, the average of their mocks has an improved correlation of 0.884 with actual draft rank. I restrict my analysis to the 91.3% of my sample (648 players) who were mocked to be drafted to ensure high-quality projections. I then calculate players' **rank shock** as his average mock rank minus his actual draft rank, $x^e_{avg} - x^a_i$, such that players who place lower than their mock experience a negative rank shock. For example, if a player is mocked to be chosen 15$^{\text{th}}$ but is actually picked lower, say 20$^{\text{th}}$, this constitutes a rank shock of -5. Figure 2 displays the distribution of players over a domain of positive and negative rank shocks.

3.2 Salaries

Player salaries reflect their aggregate income from playing professional basketball in the first five years following their NBA draft. I focus on the first five years as a player's draft outcome determines his place in the NBA rookie scale, which lays out soft guidelines for salaries in the first three years, a percentage increase for the fourth year, and another

[13]Player affiliations are usually an American college, but could also be an American high school or an international basketball team.

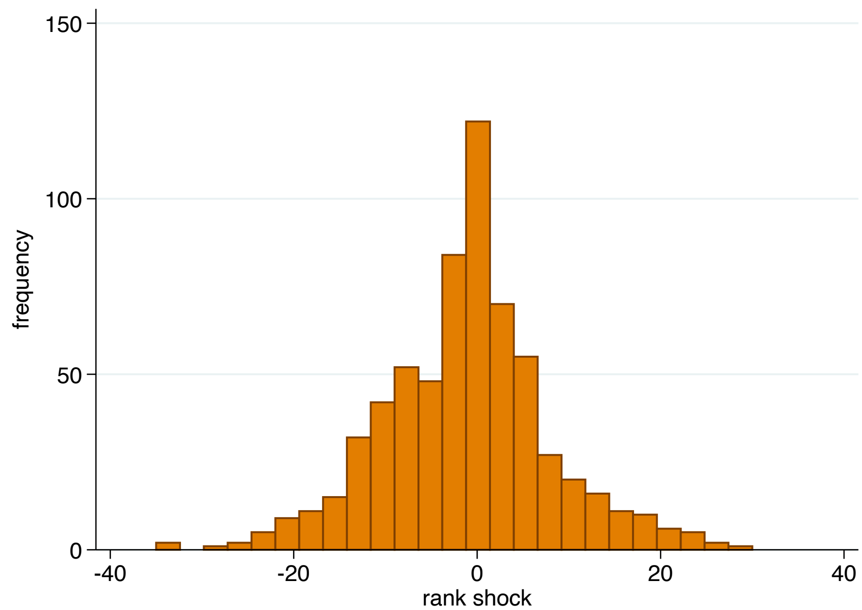


Figure 2: **Distribution of rank shocks.** A player's rank shock is calculated as his predicted (mock) rank minus his actual draft rank.

percentage increase for a qualifying offer in the fifth year. For comparability across cohorts, I convert all contract values to (after-tax) 2020 US dollars.[14]

Annual salary contracts fall into one of three categories: NBA (major league), NBA minor league,[15] or overseas (namely Europe, Asia, Australia, and South America). To construct my salaries database, I compile a chronological list of the team(s) each player signs a contract with, grouped by year, by cross referencing players' Wikipedia page, proballers.com, and basketball-reference.com. Any missed year without a signed contract is assigned a value of zero.

A database for NBA (major league) salaries can be found on Hoops Hype.[16] Minor league salaries before 2017 fell into one of three tiers, A, B, or C, which paid a flat base

[14]I convert individual year-t salaries into year-2020 salaries by multiplying the realized salaries by an NBA salary cap growth factor, which I calculate as 2020's salary cap divided by year t's salary cap. This accounts for both inflation and real growth in the payroll budget.

[15]The NBA minor league is also known as the G League or the development league.

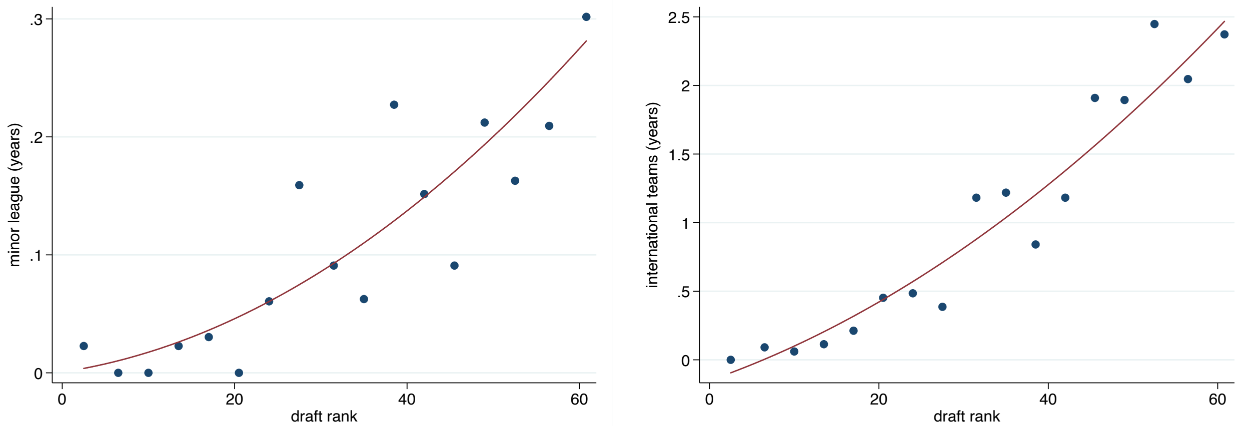
[16]While many sites track NBA salaries, such as ESPN, I find the Hoops Hype database to be most suitable for my purposes. The ESPN database contains some errors, where a few players are listed as having a salary 1 year before their draft of the same amount as their first-year NBA salary. It is also incomplete, especially for salaries before 2005. For example, although LeBron James was drafted in 2003, his salary data is only available starting in 2005, despite the fact that his 2003 contract was publicly disclosed on various sports articles). Finally, unlike Hoops Hype, ESPN salary figures do not seem to be updated throughout the season, which is especially crucial for my purposes in the case of non-guaranteed or partially guaranteed early termination contracts.

salary (around $25,000, $19,000, or $13,000, respectively) to all players belonging to the same tier. Given that the player ratio is 2:3:5 across tiers (i.e., tier A comprises 20% of all G League players), I assume that all the players in my sample who are playing in the minor league are in the top 20% and receive tier-A salaries. I track tier-A salaries as they appear in Business Insider and Sports Illustrated articles and athlete blogs, which increases only slightly over the years. All call-up contracts of minor league players to the major league, such as 10-day contracts with NBA teams, are prorated (i.e. for 10 out of the total number of days in the NBA season) according to the NBA Collective Bargaining Agreement. The base salary figure for these call-ups is given by the NBA minimum salary scale, which provides different minimum salaries for each NBA season, depending on previous years of experience in the NBA.

Overseas contracts, however, require individual data collection for each player-year-team observation, as there are no centralized salaries databases. I begin with collecting known contracts, usually from media coverage of superstar contract deals or Europe's highest paid players rankings, but also from legal disclosures of contractual dispute cases brought before the Basketball Arbitral Tribunal. For the remainder, I estimate player salary using team payroll budget and team roster size, weighting players by their average minutes played per game. This provides a more accurate salary estimate than the standard approximation of dividing team budget equally between the players, as NBA draft picks playing overseas are often above-average or top players on these foreign teams, earning well above the average contract. I automate this process for consistency by collecting further information on contract terms (i.e. whether the contract is for the season or renewable monthly), contract start/end date, season start/end dates, and make note of whether any early drop out was due to injury, poor team fit, a mutual release agreement, or being fired. See Appendix A for details.

Despite the fact that minor league and overseas contracts are not technically part of the NBA rookie scale, they are a realistic outcome of lower draft picks. Figure 3 shows the number of years (out of the initial five years post-draft) that players spend in the minor league and internationally by draft pick in equal-sized bins. Similar to how the salary expectations curve smoothly decreases by draft rank, the number of years players can expect to spend playing professional basketball outside of the major league is also smoothly increasing by draft rank. Thus, minor league and international salaries enter into player salary expectations by draft rank.

While US basketball contracts for the NBA and G League are expressed as before tax, European contract figures are after tax and denominated in Euros. To calculate after-tax salaries in US dollars, I convert Euros to US dollars using historical annual average closing exchange rates, and impose an effective tax rate of 41% on US contract



(a) (b)

Figure 3: The subplots show the number of years (out of the initial five years post-draft) players spend (a) in the minor league and (b) with international teams by draft pick in equal-sized bins.

values.[17] Figure 4 displays the distribution of players' initial five-year after-tax salaries.

3.3 Family & Demographic Info

For each player, I collect family data on engagement, marriage, and baby birth years – indexed by partner id – as well as relationship characteristics, such as the couple's age gap and whether they started dating before the NBA draft. To do this, I draw upon an assortment of primary sources, such as interviews, player bios, social media accounts,[18] police reports, and legal documents. I also collect information on various demographic variables, such as players' birth year, ethnicity, nationality, educational attainment, college major, and post-retirement occupation, as well as their partner's

[17]I obtain my ballpark effective tax rate based on ESPN (2017)'s sports tax expert's estimates. While federal, state, and municipal tax rates vary by state, 41% is the average. Note that while more accurate tax estimates could be obtained from, e.g. NBER taxsim, we are trying to estimate the expected salaries curve E[y|x] from the point of view of NBA draft hopefuls. While players and their coach/agent may have the experience to estimate ballpark contract values, it is highly unlikely they are looking up thousands of state and municipal tax rates, not to mention jock taxes for each out-of-state game, and accounting for marital status, family tax deductions, etc.

[18]Within social media platforms, I focus primarily on Instagram or Facebook. Authenticity of public figure accounts is generally social-network verified, which is indicated by a blue checkmark next to the account name. In absence of this, I verify authenticity by checking the accuracy of the account's personal details, and importantly, that the account's followers / network connections include the player's family and fellow teammates so as to avoid not only mistaken identity but also impostor accounts. Fan pages and forums are also consulted in the data collection process, but only insofar as the claims can be cross-checked with primary sources such as verified social media accounts or archived content.

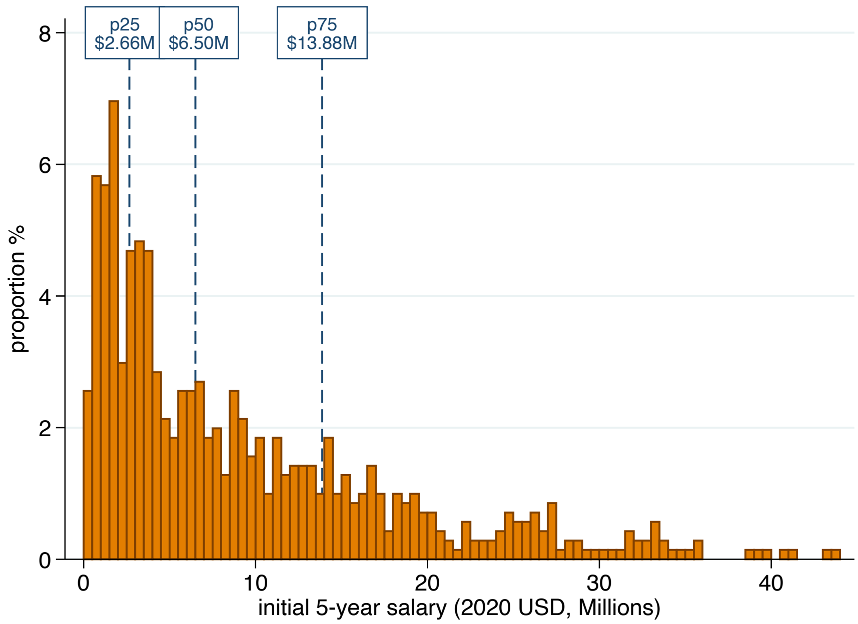


Figure 4: **Distribution of salaries.** This histogram shows the initial five-year after-tax salaries of players in the 2003-2013 draft cohorts, denominated in 2020 US dollars; p25, p50, and p75 represent the first, second, and third quartiles, respectively.

birth year, education, and occupation.[19] Note that the data collection process does not rely on wikis due to their general absence of detailed or up-to-date family information.[20]

My family data collection methodology proceeds as follows. I begin by searching for precise dates regarding marriage proposals, weddings,[21] and baby births in primary sources. Sources can take the form of news or social media announcements, time-stamped celebratory pictures or status updates, player bios, personal blogs, wedding websites, and baby/wedding registries. Alternatives include time-stamped throwback photos, and anniversary and birthday celebrations.[22] Additionally, sports interviews often include questions about the players' significant other and family life. With successive interviews specifying girlfriend, fiancée, or wife and the ages of the children, the relevant years can

[19]Partner birth year can be found on online public records databases, such as FamilyTreeNow, SearchQuarry, VoterRecords, and WhitePages; common knowledge on wives/girlfriends and relationship background is available on PlayerWives.

[20]While all draft picks have a Wikipedia page, only a handful of superstars have a 'Personal life' section with information on wife and children. Additionally, celebrity wikis such as Playerswiki and DatingCelebs often display an out-of-date "single" relationship status and an incomplete list of children.

[21]On the rare occasion where a player signs marriage papers and enters into a civil marriage in a year that is different from the wedding ceremony, I record the marriage year based on the earlier of the two dates.

[22]In absence of informative photo captions on social media, the anniversary or birthday being celebrated can often be found in the photo hashtag, a clue in the photo such as a throwback time stamp or the writing on the cake, comments from friends/family, or a reply to a fan's question.

be backed out given article publication dates. This technique can likewise be applied to police reports and legal documents, such as for domestic disputes, child support, and divorce proceedings, which state the legal relationship between relevant parties and the ages of any minors. For the roughly 20% of players who go on to play primarily overseas, a main resource apart from social media is the local media of the country in which the player is active and/or the native country of an international player. To extract these data, I employ keyword search in the country's respective language, including but not limited to Spanish, Greek, Hebrew, Croatian, Russian, Chinese, and Korean, and then translate the source language back to English using Google Translate. Appendix B offers detailed examples on how resources can be located when a direct search yields no hits, and highlights some potential pitfalls of automating the data collection process to bear in mind.

I generate an *ever-committed* relationship indicator that assigns a value of 1 to player-partner pairs who have ever been engaged and/or married within a defined post-draft observation window and a value of 0 otherwise. A value of 0 encompasses not only unattached singles but also any relationship where an engagement or marriage does not occur within the observation window. For comparability across cohorts, I set the default time span for observing marriage outcomes as five years to correspond with the initial five-year salaries. The main outcome variable of interest, the *ever-married* relationship indicator, is defined similarly. Marriage year is unknown for 8.87% of the entire sample and for 6.03% of drafted players. However, there does not appear to be selection on players with missing marriage information. T-tests reveal they do not have meaningfully different individual characteristics, such as age at draft, height, draft rank, college attainment, etc. To be conservative, I assign players with unknown marriage year an ever-married value of zero. Alternatively, excluding these players does not meaningfully change my results. With marriage year(s), I can additionally calculate *age at first marriage* as the difference between marriage year and birth year.

Figure 5 illustrates the percent of players who married by a certain age for each draft cohort. There are no clear shifts in the curves over 2004-2013 in the sense that more recent draft cohorts are not seen marry later or at lower rates than earlier cohorts. The ever-married rates by age 35 for the sample are roughly 10% higher than that of general population men in the U.S.

Additionally, I am able to precisely link children to player-partner pairs, and exclude any step-children from a partner's previous relationship. Similar to the ever-committed and ever-married indicators, I generate an *ever-father* indicator for whether players have children within the first five years of the draft. I also count the *number of children* a player has within the post-draft observation window.

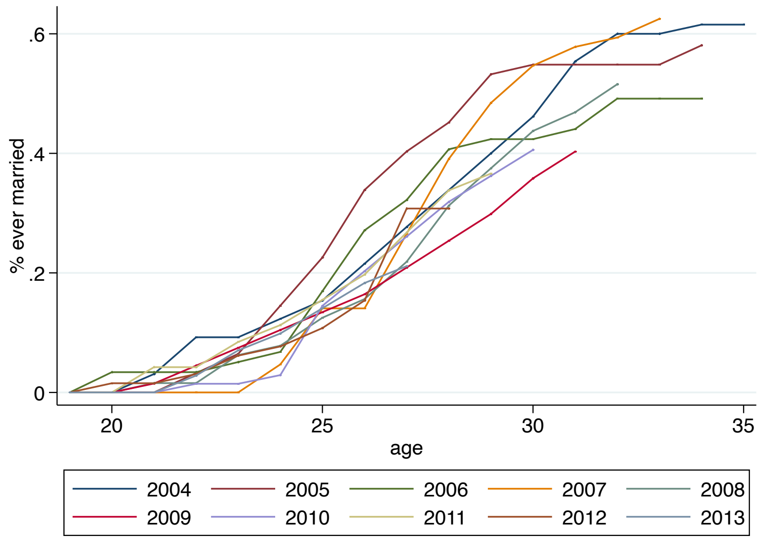


Figure 5: **Ever-married rates by draft cohort.** Each curve represents the percent of players who married by a certain age for a given draft cohort.

3.3.1 Accuracy

This dataset avoids the speculative nature of attempting to track entire dating histories by focusing on committed relationships, for which there is reliable documentation.

One potential concern is differentiating between those who are single and those who are in a committed relationship but attempt to keep their personal life private. This ambiguity arises with players who are active on social media but whose posts may be solely about basketball and are devoid of family or life-related activities. Players who fall into this category could be unattached (but may not publicly confirm they are single), privately in an uncommitted relationship, or potentially, privately in a committed relationship. First, recall that our commitment identifier does not require us to distinguish between players who are privately dating from those who are single. This simplifies matters substantially as many players avoid posting or talking about pre-commitment girlfriends. This being said, however, the identity or existence of these girlfriends is often known, usually through the girlfriend's own social media. Furthermore, this privacy does not extend to fiancées and wives to the same extent. Privacy in committed relationships generally takes the form of having a secret wedding with only close relatives and friends, keeping the partner's social media account hidden from the public, or not disclosing her full name or the names of their children in an interview. With public figures, however, private ceremonies are reported on after the fact, and the existence of a fiancée, wife, and children is acknowledged during interviews.

Thus, this concern is unlikely to affect the results.

4 Methodology

To test the causal effect of salary on propensity to marry, I exploit unanticipated changes to player drat order as an IV that exogenously shifts player salary. While we would ideally like to regress marriage outcome directly on salary, the coefficient would be biased away from the true causal effect, as salary is endogenous, due to both omitted variable bias and reverse causality.

To see how rank shock affects player salary, let us consider the intuition from the player's perspective. Recall that these players are top college or top international players. Leading up to draft night, they know their game performance, their coach or agent will have given them a confidence interval of where they can expect to place in the draft, and they will have attended the NBA *draft combine*, which is a multi-day event where shortlisted NBA hopefuls are interviewed, measured, take medical tests, and perform drills for an audience of NBA coaches, general managers, and scouts. The expert analysts also have access to this information and more: not only can they access game stats and video footage online, but they also attend the multi-day event to spectate and interview the players, players' agents/coaches, and NBA personnel. Thus, the expert analysts can give each player a bigger picture sense of their competitiveness relative to the other prospective players. Combining this with some private information from their insider sources within the NBA teams, the experts distill the available information into an unbiased point estimate for each prospective player, which we call his mock rank x_i^e. This corresponds to a certain initial contract size and corresponding five-year salary expectation.

Draft night, on the other hand, is more of a black box to players. This is because, outside of the top three players, even NBA teams themselves cannot know for sure who they will pick as they walk into draft night. While teams may have their eye on a certain player, they cannot perfectly anticipate how other teams will tie-break between players. Therefore, there is no guarantee that the player they are considering will still be available by their turn to draft or if an even better player the team did not anticipate being able to acquire is still available (see Section 2 for examples). Additionally, teams must make their decision within minutes.[23] All of this contributes

[23] NBA teams are given 5 minutes on the clock to select a top 30 draft pick and 2 minutes for a 30-60 draft pick, with a 30 second warning at the 4:30 and 1:50 minute marks, respectively. Exceeding this time limit incurs substantial penalties from the NBA League Office, including fines, loss of future draft choices, or suspension of relevant personnel.

to the randomness of draft night, upon which players realize their actual draft rank x_i^a. If x_i^a is different from x_i^e, the player suddenly faces a different initial contract and five-year salary expectations. Thus, we can think of a player i's *rank shock* $x_i^e - x_i^a$ as an unanticipated change in his draft order that generates an exogenous income shock. This intuition is summarized in Figure 6.

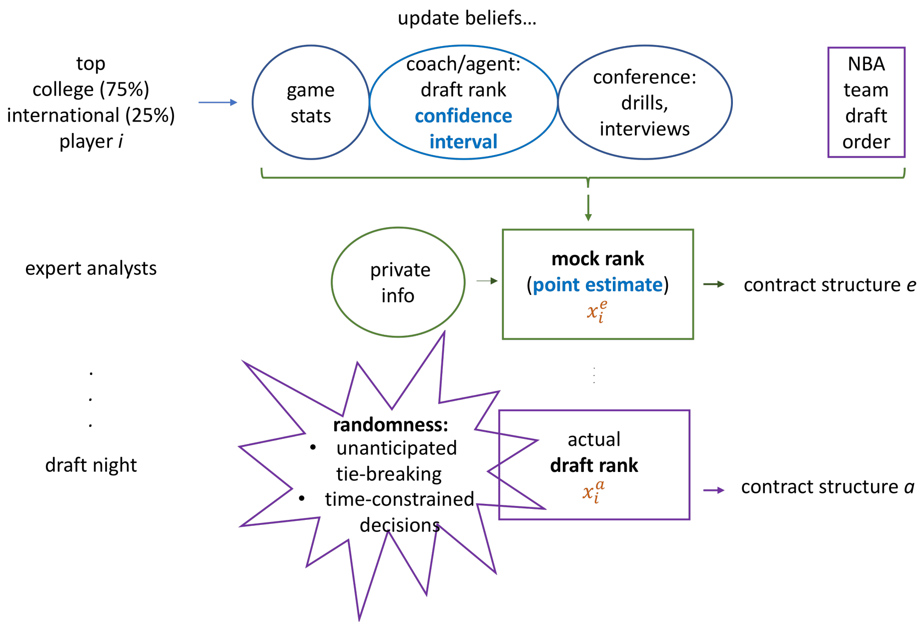


Figure 6: **Natural experiment setup.**

4.1 Identifying Assumptions

To use rank shock as an IV for salary, it needs to satisfy three assumptions:

Assumption 1: Strong first stage. Rank shock should have a strong positive relationship with salary. We can test for this in a first-stage regression of log salary over rank shock and any other second-stage controls. This assumption is satisfied if the rank shock coefficient is positive and the F-statistic is greater than 10.

Assumption 2: Independence. Rank shock should be randomly assigned on both players' earning potential and marriageability characteristics. In theory, the high-quality draft predictions from well-respected experts coupled with draft night randomness resulting from team tie-breaking between players should result in random assignment and unanticipated rank shock. We can also test for random assignment of rank shock on *observable* characteristics. For example, earnings potential could be proxied by the expected salary associated with player mock rank, which is given by the fitted log-quartic curve in Figure 1. Observable marriageability variables include age at draft, height, years of college, an indicator variable for college graduation, and college (NCAA) ranking. As well, we can check for the existence of idiosyncratic player-team match quality that may not be known to the experts. That is, when an NBA team drafts a player ahead of their mock, is there a synergistic effect between the player and the team, such as great sportsmanship or coachability that the experts did not take into account? To test this, we can see if players who get positive rank shocks tend to play more games per season, minutes per game, or are retained for more years by their initial drafting team. If this is true, then players should be able to anticipate at least part of the rank shock, in which case it would not be an unanticipated shock. Additionally, we can test the quality of the mock drafts by checking for information asymmetries between the mock drafters and the NBA teams. For example, we can compare mock and draft ranks in terms of their predictive power for realized player salary. If the information asymmetry between the experts and NBA teams is not statistically significant on average, then disparities between mock and draft ranks should be unanticipated.

Assumption 3: Exclusion restriction. Rank shock should only affect players' propensity to marry by changing the size of the initial contract that players sign. While this assumption cannot be tested, I consider potential threats to its validity. Unlike large economy-based shocks, such as with a sudden switch between employment and unemployment, one strength of this setup is that there is no simultaneous shock to time use. However, one potential violation of the exclusion restriction is that, given a certain mock rank, a different rank shock means the player would be assigned to a different NBA team and therefore go to a different location. However, we can easily internalize this in the first stage. As we will see, the results remain robust for two reasons. First, basketball players travel extensively and play half of their games away home court regardless of their placement and so are not constrained to their home team location. Second, by virtue of the random assignment of rank shock, there is also no selection on location. Results that remain robust to the inclusion of further controls is an encouraging sign that the exclusion restriction is not violated.

Satisfying these three assumptions, we can proceed to estimate the following probit two-stage least squares:

$$\Phi(Y_i) = \beta_1 ln(salary_i) + \beta_2 mock_i + X_i'\beta_3 + \mu_{nat} + \mu_{ethn} + \mu_{loc} + \mu_{coh} + \epsilon_i, \quad (1)$$

with the following first stage:

$$ln(salary_i) = \lambda_1 rankshock_i + \lambda_2 mock_i + X_i'\lambda_3 + \mu_{nat} + \mu_{ethn} + \mu_{loc} + \mu_{coh} + \nu_i, \quad (2)$$

where players are indexed by i, Y_i is our marriage outcome of interest (e.g. an ever-married indicator that takes a value of 1 if player i is married within the five-year post-draft window and 0 otherwise), $ln(salary_i)$ is log realized salary, $mock_i$ is average expert projection, X_i is a vector of individual characteristics (age at draft, age squared, and height),[24] and μ_{nat}, μ_{ethn}, μ_{loc}, and μ_{coh} are nationality, ethnicity, location, and draft cohort fixed effects, respectively.

5 Main Results

5.1 Naïve Regression

Let us begin with a naïve specification, where we regress marital status directly on salary. Table 1 displays the regression results without and with controls in columns (1) and (1'), respectively. For comparability with the main 2SLS results, the controls in naïve regression (1') will be the standard individual characteristics, mock rank, age at draft, age squared, and height, and fixed effects for nationality, race, location, and draft year.

Table 1: Naïve Effect of Salary on Marriage Likelihood

	(1) marry within 5 yrs of draft	(1') marry within 5 yrs of draft
log salary	0.016	0.032
	(0.013)	(0.020)
controls	N	Y
N	635	600

Notes: Naïve regressions of marital status over salary with and without controls. Controls include mock rank, age at draft, age squared, height, nationality FE, race FE, location FE, and draft year FE. Standard errors are given within parentheses. Regressions exclude players who married before the NBA draft. Levels of significance: * 10%, ** 5%, and *** 1% level.

[24] I exclude $education_i$ from X_i for two main reasons. First, education is not statistically significant in my regressions. Second, whether a player graduates from college is almost entirely a mechanical result in the case of NBA players: prospective draft picks generally leave college in the year in which they are ready for the NBA draft. The opportunity cost of finishing college when they could be earning an NBA salary is high, especially if they do not plan to use their degree.

The our the coefficient of interest doubles moving from (1) to (1') illustrates an endogeneity problem. While not statistically significant, the underlying causal effect could still be positive or negative. For one, salary could be correlated with any number of variables in the error term that are also correlated with marriage outcome, which biases our naiv̈e estimate away from the true causal effect. Common examples of omitted variables that may upward bias our estimate are ability and interpersonal skills, since we may be falsely attributing the effects of these attractive qualities to salary by not including them in the regression. Conversely, a source of negative bias could be greater fame as a result of higher ability, rendering the player more selective about the partner he chooses to commit to. While controlling for more individual characteristics and various fixed effects may reduce some of the bias, unobservables and other variables that are difficult to measure limit us in our ability to internalize all the relevant terms to fully resolve the endogeneity problem. Additionally, there could be reverse causality in that, it may not be salary that is inducing marriage but instead that married players earn higher income as a result of being in more stable relationships.

A solution to these problems is to find a randomly assigned shock that exogenously shifts salary before marriage decisions are made to use as an IV. In the case of the NBA drafts, rank shock fits such a description. Before we dive into the two-stage least squares (2SLS) regression, let us first check that rank shock satisfies a strong first stage, independence, and the exclusion restriction.

Table 2 displays the the first-stage results of regressing salary on rank shock, with the progressive addition of controls. As we can see, rank shock has a strong and robust relationship with salary across the board. The F-stat is above the rule of thumb of 10, even in specification (1) without controls, and the relationship between rank shock and salary remains unchanged, regardless of controls. The controls in specification (1') and (1"), however, provide the added benefit of reducing the unexplained variation in salary, which allows us to obtain more precise estimates of the coefficient of interest, as evinced by smaller standard errors and larger F-statistics. From this exercise, we learn that being drafted earlier by one spot increases initial five-year salary by 1.9%.

Next, let us check that rank shock is randomly assigned on marriageability characteristics and earning potential. In addition to the intuition that team tie-breaking between players leads to unanticipated changes in draft order (see Section 4), let us also check that the relationship between rank shock and *observable* characteristics is not statistically significant. To do this, rank shock can be separately regressed over the various standardized characteristics. Figure 7 illustrates the results of these tests. As we can see, rank shock is randomly assigned on observable individual characteristics, such as age at draft, height, years spent in college, whether the player graduated from college, and the

Table 2: 2SLS – First Stage

	(1) log salary	(1') log salary	(1") log salary
rank shock	0.019***	0.019***	0.019***
	(0.005)	(0.004)	(0.004)
mock	N	Y	Y
location FE	N	Y	Y
other controls	N	N	Y
N	635	634	633
F-stat	13.54	24.01	26.58

Notes: The first stage displays the strength of the relationship between rank shock and salary, with and without controls. Other controls, besides mock rank and location FE, include age at draft, age squared, nationality FE, race FE, and draft year FE. Standard errors are in parentheses. Regressions exclude players who married before the NBA draft. Levels of significance: * 10%, ** 5%, and *** 1% level.

ranking of the college's basketball program (according to the NCAA poll of top college coaches) in the season leading up to the draft. Equally importantly, rank shock is also randomly assigned on earnings potential, which we can estimate with the expected salary associated with mock rank. Figure 8 further shows that rank shock is not only uncorrelated with earnings potential on average but is also randomly assigned all along the mock ranks, which suggests the mocks are high quality and unbiased.[25]

Returning to Figure 7's post-draft outcomes, we see that NBA teams who draft players earlier than projected by experts do not play them in more games per season. If anything, the teams give these players weakly fewer minutes per game and retains them for weakly fewer seasons. Therefore, it does not seem likely that NBA teams are drafting players ahead of their mock due to privately observed idiosyncratic player-team match quality.

Additionally, we can check whether there is information asymmetry between the expert mock drafters and NBA teams by comparing the predictive power of mock rank versus draft rank on actual realized salary. To do this, we can calculate the correlation between expected salary given the mock and realized salary, corr(log E[y|mock], log salary)=0.7096. In comparison, the correlation between expected salary given the draft and realized salary is corr(log E[y|draft], log salary)=0.7112. Testing whether the correlations are equal yields a p-value of 89.3%. Thus, we cannot reject the null hypothesis. As mock and draft rank are equally predictive of actual salary, there does not seem to be information asymmetries between the expert mockers and NBA teams

[25] The diagonal on the lower right is a purely mechanical result: Players who are mocked to be picked towards the end have little potential for a negative shock.

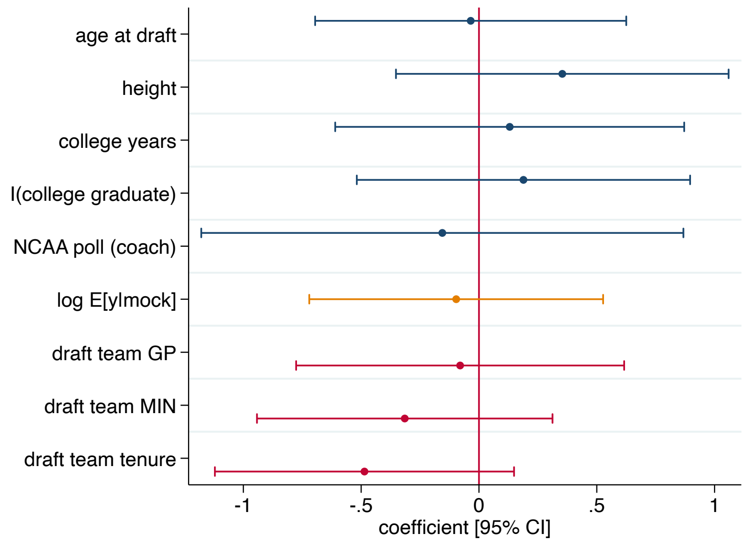


Figure 7: **Random assignment of treatment**. This plot shows the coefficient and 95% confidence interval of the relationship between rank shock and various (standardized) individual characteristics (blue) and earning ability (orange). The variables include age at time of draft, height, years spent in college, an indicator for college graduation, college basketball ranking (from the NCAA coach poll) in the year leading up to the draft, and expected salary given mock rank. Additional tests include three post-draft outcomes with the initial drafting team (red): average games played per season, average minutes played per game, and years of tenure.

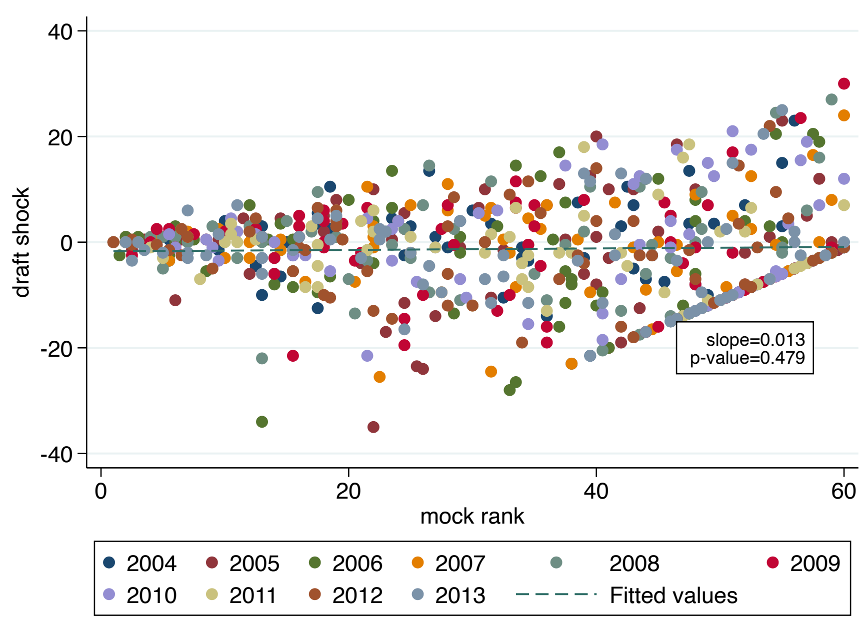


Figure 8: **Rank shock variation.** This graph illustrates the variation in rank shock at every level of the predicted (mock) ranks.

on average.

Finally, we need rank shock to only affect marriage outcome through our endogenous variable, salary. Intuitively, players who receive a rank shock are simply assigned to a different initial contract than the one they expected, but otherwise all become professional basketball players. While rank shock also affects which team the player joins and therefore also determines his location, we will see in the second stage that controlling for mock rank and location fixed effects does not meaningfully alter the results. Indeed, the result is also robust to controlling for the remainder of the covariates. Additionally, to show that the relationship between male earnings and marriage rates is positive, i.e. to rule out the possibility that marriage is an inferior good, any exclusion restriction violations that negatively bias our results do not harm the story, as it would simply mean that our estimates serve as a lower bound for the true causal effect. Only violations that positively bias our results may be concerning. To the extent that we believe being drafted earlier is associated with a marginal fame shock, I argue this is likely a negative-bias story, as the naïve regression lacks statistical significance. Given that ability and fame are two major sources of endogeneity for salary, and ability is generally a source of positive bias, fame and selectivity of partner as a result of fame seems like a likely source of negative bias. A potential check for this is the external validity check of Section 7.3. Dropping the top draft picks, where the ability and fame differential are largest, I see that the causal effect, if anything, only becomes more positive and significant.

If the above assumptions hold, then rank shock is a valid IV for salary, and the second stage of the probit 2SLS[26] yields average causal effects. Table 3 displays the results for all three specifications. The causal effect of salary on propensity to marry does not change meaningfully across specifications, which is an encouraging sign that the exclusion restriction is not violated. According to specification (2”), a 10% increase in salary increases propensity to marry by 1.73 percentage points. Given the baseline marriage rate of 19.4%, this translates to an increase of 8.9 percent. Regardless of the specification, I find that marriage is indeed a normal good for the 2004-2013 NBA draft cohorts.

[26] I employ probit 2SLS instead of a linear probability model (LPM), as LPM does not fit the data well, with 13.6% of the predicted values falling outside of $[0, 1]$.

Table 3: Probit 2SLS – Second Stage: Propensity to Marry

	(2) marry within 5 yrs of draft	(2') marry within 5 yrs of draft	(2") marry within 5 yrs of draft
log salary	0.144**	0.161**	0.173**
	(0.064)	(0.074)	(0.071)
mock	N	Y	Y
location FE	N	Y	Y
other controls	N	N	Y
N	635	606	600

Notes: The second stage shows the causal effect of salary on propensity to marry, with and without controls. Standard errors are given within parentheses. Controls correspond to those of the first stage: Other controls, besides mock rank and location FE, include age at draft, age squared, nationality FE, race FE, and draft year FE. Regressions exclude players who married before the NBA draft. Levels of significance: * 10%, ** 5%, and *** 1% level.

6 Alternative Specification

Alternatively, I can also regress marriage outcomes directly on the exogenous draft income shock. In this case, only random assignment is necessary for a causal interpretation. To find the draft income shock, let us revisit the expected salaries curve and an example in Figure 9.

Suppose a player is projected to be picked 12$^{\text{th}}$ but is actually drafted 40$^{\text{th}}$. Both of these rankings can be mapped to expected salaries using the log-quartic curve of best fit from Figure 1. By virtue of the natural experiment surrounding NBA drafts, we can decompose each player's realized salary y_i into three components:

$$y_i = E[y|x_i^e] + \underbrace{E[y|x_i^a] - E[y|x_i^e]}_{\text{Draft income shock}} + \underbrace{y_i - E[y|x_i^a]}_{\text{Idiosyncratic component}}, \tag{3}$$

where his base salary is the expected salary given his mock $E[y|x_i^e]$, his exogenous draft income shock $E[y|x_i^a]-E[y|x_i^e]$ is the difference between the expected salaries associated with his actual draft rank and his mock; and his idiosyncratic component $y_i - E[y|x_i^a]$ represents the extent to which his realized salary y_i deviates from the expected salary given his draft rank. Thus, in place of rank shock, we can use the exogenous **draft (income) shock** $E[y|x_i^a] - E[y|x_i^e]$ directly in the alternativef specification.

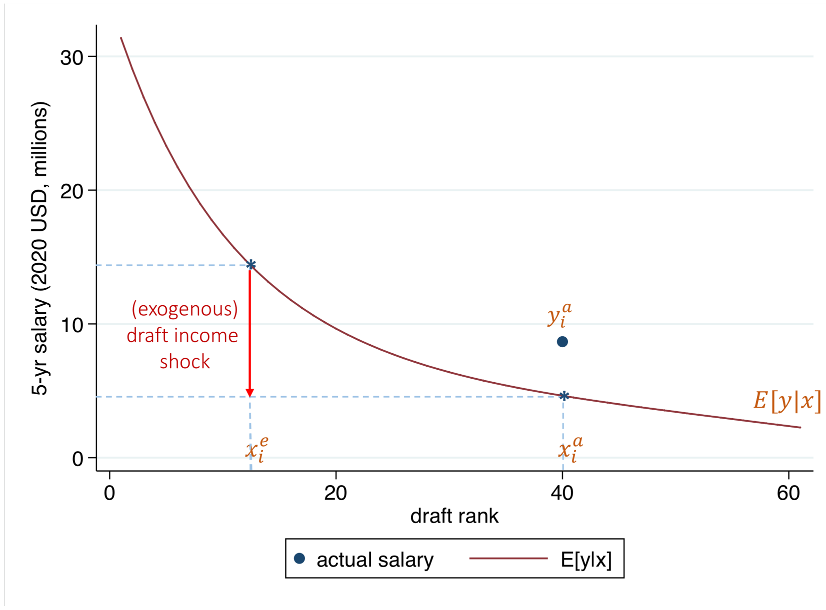


Figure 9: **Salary components.** The $E[y|x]$ curve represents expected player salaries by draft rank. Projected (mock) rank x_i^e and actual draft rank x_i^a are associated with salary expectations $E[y|x_i^e]$ and $E[y|x_i^a]$, respectively. Thus, a player's actual realized salary y_i^a can be decomposed into his expected salary given his mock ($E[y|x_i^e]$), exogenous draft income shock ($E[y|x_i^a] - E[y|x_i^e]$), and idiosyncratic component ($y_i - E[y|x_i^a]$).

6.1 Regression Setup

In the alternative specification, we can regress our marriage variable of interest directly on draft (income) shock $E[y|x_i^a] - E[y|x_i^e]$ in place of salary as follows:

$$Y_i = \beta ln(draftshock_i) + \delta mock_i + X_i'\gamma + \mu_{nat} + \mu_{ethn} + \mu_{loc} + \mu_{coh} + \epsilon_i, \qquad (4)$$

where Y_i represents any marriage-related outcome variable for player i, our variable of interest $ln(draftshock_i)$ is the natural log of the exogenous draft income shock, i.e. $ln(E[y|x_i^a]/E[y|x_i^e])$, with the standard controls. Essentially, this setup asks: For players of the same mock rank, how does an exogenous draft income shock affect their marriage outcomes, controlling for individual characteristics? Table 4 displays the results of the alternative specification, with and without controls.

Note that we cannot directly compare the alternative specification results with those of the 2SLS, as we are measuring different things. While the 2SLS shows the effect of a change in *realized* salary on marriage, the alternative specification gives us the effect of a change in *expected* income on marriage. To compare the two, let us next run a "first stage" to see how an *expected* draft income shock translates into salary. Table 5 displays the results.

Table 4: Alternative Specification: Propensity to Marry

	(3) marry within 5 yrs of draft	(3') marry within 5 yrs of draft	(3") marry within 5 yrs of draft
log draftshock	0.072*	0.085*	0.093**
	(0.043)	(0.045)	(0.042)
mock	N	Y	Y
location FE	N	Y	Y
other controls	N	N	Y
N	637	606	600

Notes: The alternative specification shows the causal effect of the exogenous draft income shock on propensity to marry, with and without controls. Standard errors are given within parentheses. Controls correspond to those of the first stage: Other controls, besides mock rank and location FE, include age at draft, age squared, nationality FE, race FE, and draft year FE. Regressions exclude players who married before the NBA draft. Levels of significance: * 10%, ** 5%, and *** 1% level.

Table 5: OLS – "First Stage"

	(4) log salary	(4') log salary	(4") log salary
log draftshock	0.491***	0.481***	0.478***
	(0.129)	(0.099)	(0.095)
mock	N	Y	Y
location FE	N	Y	Y
other controls	N	N	Y
N	635	634	633
F-stat	14.41	23.75	25.42

Notes: The regression tables show how changes in expected income translate into changes in salary, with and without controls. Other controls, besides mock rank and location FE, include age at draft, age squared, nationality FE, race FE, and draft year FE. Standard errors are in parentheses. Regressions exclude players who married before the NBA draft. Levels of significance: * 10%, ** 5%, and *** 1% level.

This "fist stage" shows us that the expectational draft income shock $E[y|x_i^a]-E[y|x_i^e]$ does not have a 1:1 mapping with salary.[27] Therefore, to calculate the effect of a 100% increase in actual salary on marriage, we need to divide the alternative specification result by the first stage.[28] This yields 0.147, 0.176, and 0.195, respectively, which is very comparable to the 2SLS results.

An additional takeaway from this exercise is the statistical significance of the alternative specification relative to the 2SLS. While the alternative specification in (3") has a statistical significance that is similar to the corresponding 2SLS specification, those of (3) and (3') are weaker. This is because using the shock as an IV isolates the portion of its variation that *directly* affects salary. Regressing directly on the shock is considerably more noisy, leading to larger confidence intervals and lower significance.

7 Robustness Checks

This section explores the main result more fully. First, I show that the average causal effect is representative over the domain of positive and negative income shocks. Second, I check that the average causal effect is robust to other observation window specifications for the marriage outcome. Third, I argue the case of external validity. Finally, I investigate whether the income-induced changes in marriage rates are driven by the man's demand or his partner's.

7.1 Non-parametric Effects

The rich variation in income shocks of my setting allows me to explore beyond the average causal effect. That is, we can check to see if the effect is linear and representative over the entire domain of positive and negative income shocks. In particular, we want to ensure that we have statistical significance for both positive and negative income shocks.

To do this, we can residualize the ever-married indicator by regressing it over all the other covariates besides the exogenous draft income shock. Collecting the error term, we can then plot it directly over the domain of negative and positive draft income

[27] This says that players picked at a rank x_i^a that is earlier than their mock x_i^e are not receiving the full expected salary as indicated by their draft rank, $E[y|x_i^a]$ – i.e. the idiosyncratic component tends to be negative. This does not pose a threat to identification, however, as the correlation between the idiosyncratic component and marriage likelihood is not statistically significant (p-vlaue = 0.38). Thus, not internalizing the idiosyncratic component in the alternative specification does not cause omitted variable bias.

[28] That is, if a 100% increase in expected salary (i.e. 48% increase in actual salary) causes a 9pp increase in marriage rates, the effect from a 100% increase in actual salary must be 1/0.48 times larger.

shocks. Figure 10 illustrates the result.

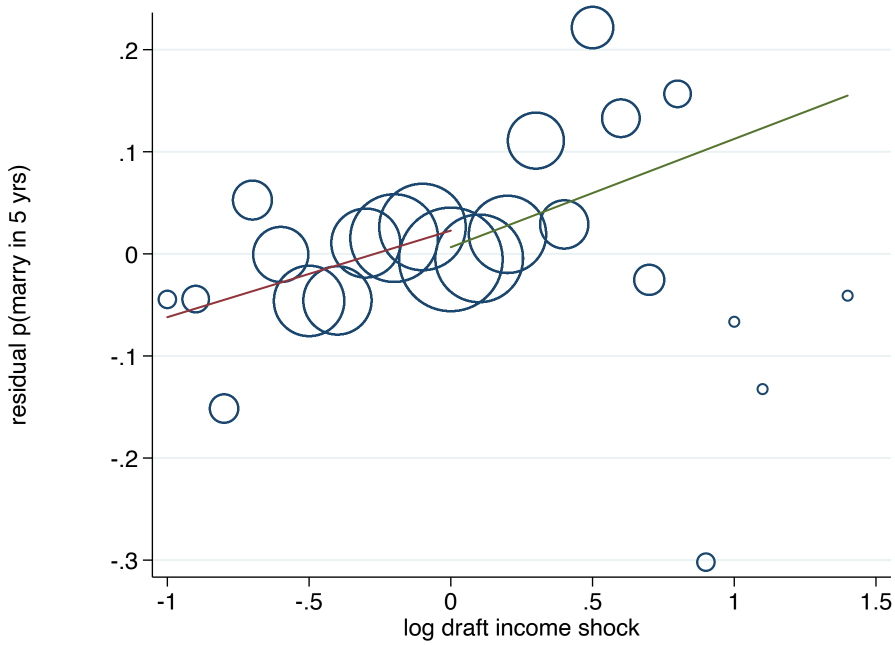


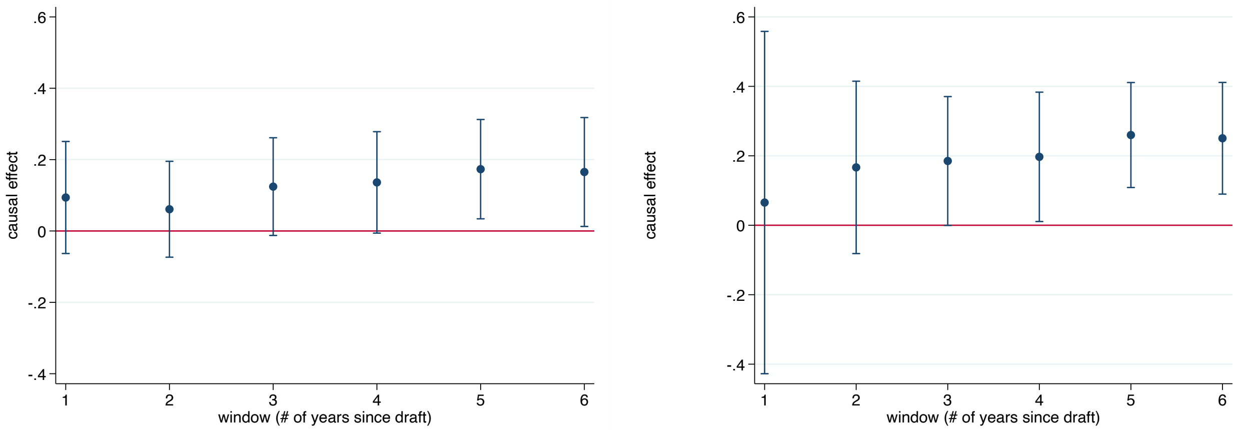
Figure 10: **Non-parametric effects of income shocks on propensity to marry.** This graph plots the residuals of the alternative specification over the domain of draft income shocks. Log draft income shock is divided into equal-sized bins. The size of each marker reflects the proportion of players in each bin. The analysis excludes players who married before the NBA draft.

The causal effect is consistently significant over both positive and negative income shocks. This confirms Autor et al. (2019)'s result for negative income shocks and addresses the literature gap for positive income shocks for men. Additionally, the causal effect appears linear over the domain of income shocks. The absence of non-linearities renders the average causal effect a representative sufficient statistic.

7.2 Varying the observation window

Recall that, for the main results, I use a five-year post-draft observation window for marriage observations to match the five-year salary window. Figure 11 illustrates how the average causal effect changes if we alter this post-draft observation window, for both the entire sample and the subset of players who attended high school and/or college in America.

Consistent with the life cycle hypothesis, the effects are positive for all observation windows. The lack of significance in the first few years of the players' career is likely due to insufficient power, given that less than 10% of player marriages take place within



(a) (b)

Figure 11: The graphs plot the causal effect of salary on propensity to marry with 95% confidence intervals for different post-draft observation windows of marriage. That is, each dot represents a separate regression coefficient for salary on ever-married within $x = 1, ..., 6$ years of the draft. Subfigure (a) is for the entire sample. Subfigure (b) is for the subsample of players who attended school in the US. Regressions exclude players who married before the NBA draft.

the first two years. With a sufficiently large sample, these too should be significant. The smaller magnitudes for these shorter observation windows are likely due to search frictions of finding a partner or that players are prioritizing their career and establishing themselves in a highly competitive industry.

7.3 External Validity

Prominent concerns against external validity typically point to basketball players' sizable NBA contracts, placing their economic status well above that of the typical man. I argue however that this concern simply suggests that my results constitute a *lower-bound* estimate for general population men. That is, if anything, men on average should be *more* likely to marry as a result of a positive income shock, not less.

First, as a within-sample illustrative exercise, let us examine how the magnitude and significance of the results change if we drop the top draft picks and progressively restrict our sample to those of lower expected salaries. Figure 12 plots the results of this sequence of regressions for propensity to marry.

For major marriage decisions, Figure 12 finds that lower expected salary players are indeed even more likely to marry due to a positive draft income shock. For comparability with the literature, let us convert these results to income elasticities of marriage demand.

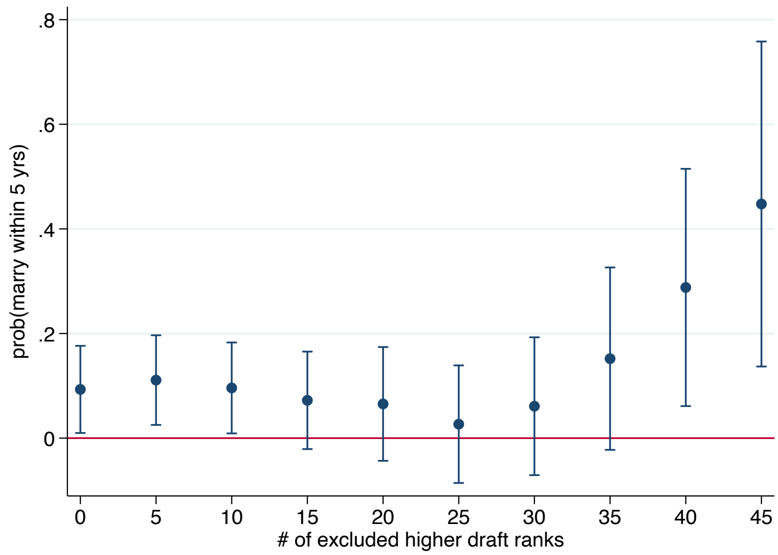


Figure 12: **Propensity to marry.** This graph illustrates the average causal effect of draft income shock on propensity to marry for 10 separate regressions: starting with the overall sample (x=0) and progressively restricting to players of lower expected salaries (i.e., excluding the top x=5, 10, 15,..., 45 draft ranks). Regressions include all controls. Confidence intervals are given at the 95% level. Regressions exclude players who married before the NBA draft.

My sample's overall elasticity[29] of 0.18 increases to 0.65 after dropping the top 45 draft picks of every cohort.[30] That the elasticity increases for lower expected salaries is consistent with Kearney and Wilson (2018)'s general population elasticity of 0.96 from the (1969-1987) Appalachian coal boom, prior to social norm shifts regarding cohabitation. While marriage is a normal good for all three (sub)samples of men, it borders on being a luxury good for general population men who, as expected, are more responsive to changes in income. Indeed, the general population figure gives support to sociologist speculation in popular media that marriage may have become a modern-day luxury good (CNBC, 2013; Forbes, 2014; The Wall Street Journal, 2020).

Intuitively, more money is more power to live life according to one's underlying preferences. If basketball players truly preferred a bachelor lifestyle, more money simply means a better ability to finance this lifestyle through the purchase of services, such as

[29]Recall the 2SLS result for the overall sample: A 100% increase in initial *five-year* earnings increased marriage rates by 89%. Thus, a 100% increase in *annual* earnings increased marriage rates by 17.8%, yielding an elasticity of 0.178.

[30]For this subgroup, a 100% increase in *five-year* income increased marriage rates by $0.448/0.653 =$ 68.55pp ($68.55/0.2098 = 327$ percent), where 44.8pp is the alternative specification result, 0.653 is the first-stage coefficient, and 20.98% is the baseline marriage rate of this subgroup. Thus, a 100% increase in *annual* earnings increased marriage rates by 65.3%, yielding an elasticity of 0.653.

housekeeping, pre-made meals, etc. In this case, we should see a decrease in propensity to marry in response to positive income shocks, but the opposite is true.

Indeed, a Pew Research Center (2019) survey of married U.S. couples finds that the reasons people marry largely fall into three broad categories: love and companionship, financial and practical considerations, and social conformity pressures.[31] At least two if not all three of these channels are substantially weaker for basketball players relative to general population men. First, their large paychecks should greatly minimize the need for financial considerations, such as income risk sharing or wealth accumulation between couples. Practical considerations, as discussed above, could also be reduced through the purchase of home production services. Second, their multitude of dating options also reduces the need for companionship through marriage. Last but not least, if society does not expect them to settle down given their abundance of wealth and dating options, then they also face minimal societal pressures to marry. Not only are their incentives to marry weaker, but they also face larger costs in the event of a divorce. This reaffirms that, if marriage remains a normal good for my sample in spite of these factors, we should only expect to see a greater response for general population men.

Extending this analysis to age at first marriage, Figure 13 finds that lower expected salary players also generally marry earlier due to a positive draft income shock. These results are directionally consistent with the seminal Becker (1973, 1974)-Keeley (1977) model, which predicts that changes in earnings that improve an individual's economic gains to marriage induce greater propensity to marry and earlier marriage (see Section 8 for a full discussion).

For major family decisions, Figures 14 and 15 replicate this analysis for propensity to have children and number of children. For major family decisions, Figure 14 finds that lower expected salary players are more likely to have children due to a positive draft income shock. Figure 15 finds that lower expected salary players generally also have more children due to a positive draft income shock. Thus, I confirm that children are also normal goods for my sample, consistent with Black et al. (2013).

7.4 Who is the income shock inducing to marry?

While propensity to marry increases in response to a positive income shock, it is not obvious whether it is the man who is being induced to marry or his partner. To investigate this, I test a tabloid theory that NBA players are vulnerable to "baby

[31] The survey of married US couples finds that love is cited by 90% of respondents as a major reason for why they decided to get married, followed by companionship (66%), wanting to make a formal commitment (63%), wanting to have children (31%), financial reasons (13%), convenience (10%), and pregnancy (6%).

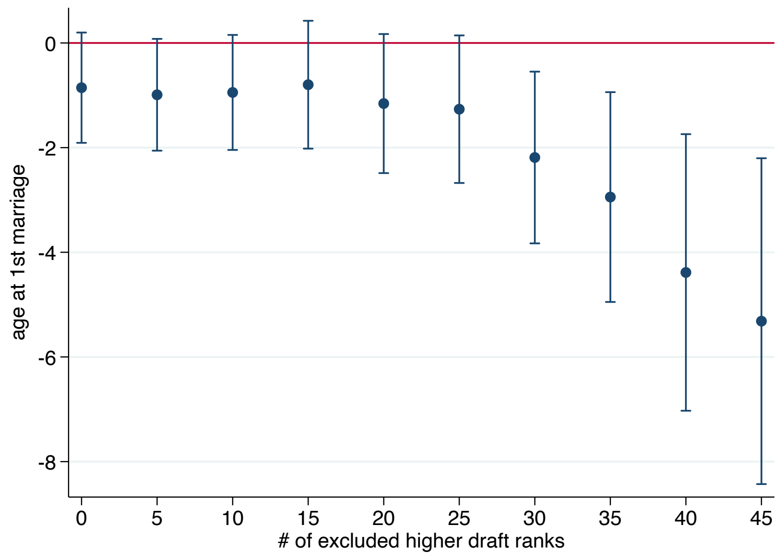


Figure 13: **Age at first marriage.** This graph illustrates the average causal effect of draft income shock on age at first marriage for 10 separate regressions: starting with the overall sample (x=0) and progressively restricting to players of lower expected salaries (i.e., excluding the top x=5, 10, 15,..., 45 draft ranks). Regressions include all controls. Confidence intervals are given at the 95% level. Regressions exclude players who married before the NBA draft.

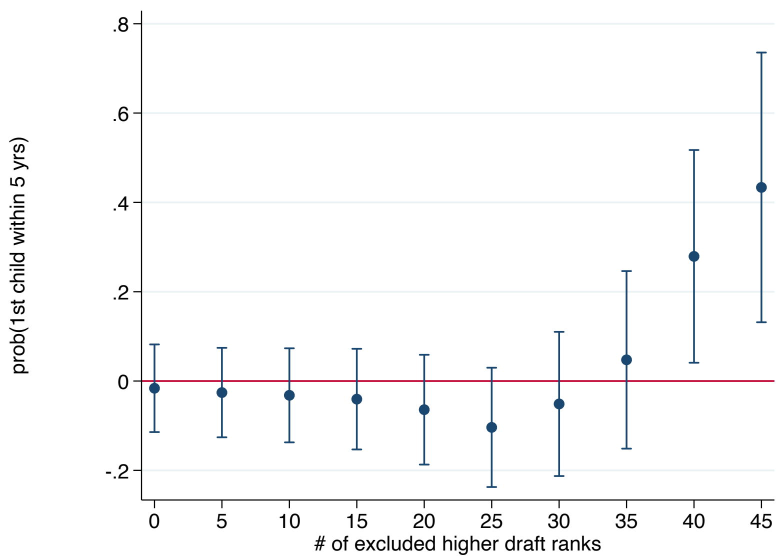


Figure 14: **Propensity to have children.** This graph illustrates the average causal effect of draft income shock on propensity to have children for 10 separate regressions: starting with the overall sample (x=0) and progressively restricting to players of lower expected salaries (i.e., excluding the top x=5, 10, 15,..., 45 draft ranks). Regressions include all controls. Confidence intervals are given at the 95% level. Regressions exclude players who had children before the NBA draft.

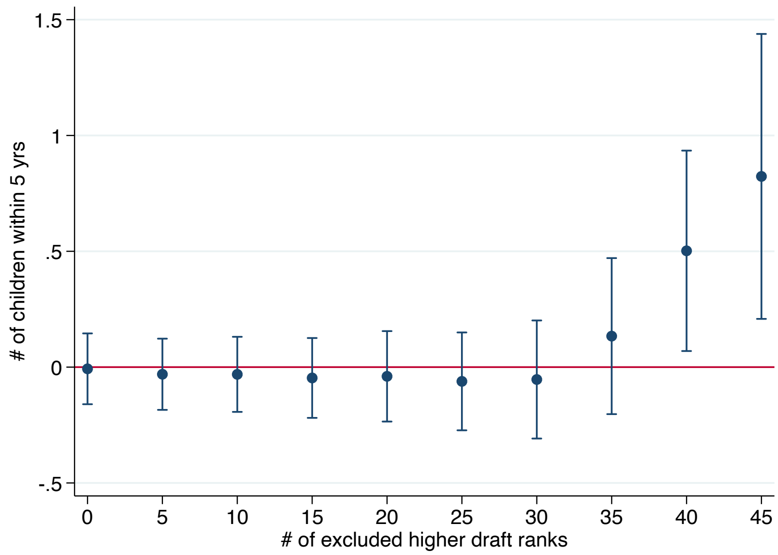


Figure 15: **Number of children.** This graph illustrates the average causal effect of draft income shock on number of children for 10 separate regressions: starting with the overall sample (x=0) and progressively restricting to players of lower expected salaries (i.e., excluding the top x=5, 10, 15,..., 45 draft ranks). Regressions include all controls. Confidence intervals are given at the 95% level. Regressions exclude players who had children before the NBA draft.

trapping" – i.e., that the partner may try to increase their chances of marriage by first having a baby with him. Figure 16, however, shows that this is not the case. While basketball players who receive beneficial draft shocks are weakly more likely to have children, they are *less* likely to have them before they are ready to commit in terms of an engagement or marriage. If we interpret a beneficial income shock as giving the basketball player more bargaining power with their partner, we see that this bargaining power allows players to better avoid the tabloid theory of being persuaded into marriage by a baby. Given that players with more bargaining power *are* more likely to commit, however, it is likely that they choose to do so in their own time and on their own terms.

This being said, knowing that a positive income shock increases marriage rates does not inform us about the quality or stability of the marriages. Figure 17 checks whether these income-induced marriages are, on average, associated with divorce and finds that this is not the case, especially for those of lower expected salary.

8 Conceptual Framework

To tie my results in with those of the empirical literature, let us consider what the economic theory predicts about the direction of the causal effect we should expect to

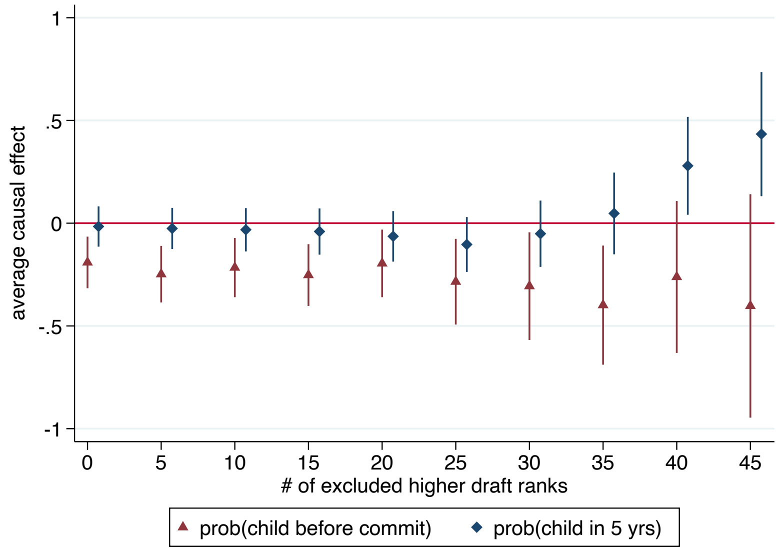


Figure 16: **Propensity to have a child versus having a child before engagement or marriage.** This graph shows the causal effect of draft income shock on propensity to have a child (blue diamonds) versus having a child *before* engagement or marriage (red triangles). Following the format of the external validity graphs, I show the effect for 10 separate regressions: starting with the overall sample (x=0) and progressively restricting to players of lower expected salaries (i.e., excluding the top x=5, 10, 15,..., 45 draft ranks). Regressions include all controls. Confidence intervals are given at the 95% level. Regressions exclude players who had children before the NBA draft.

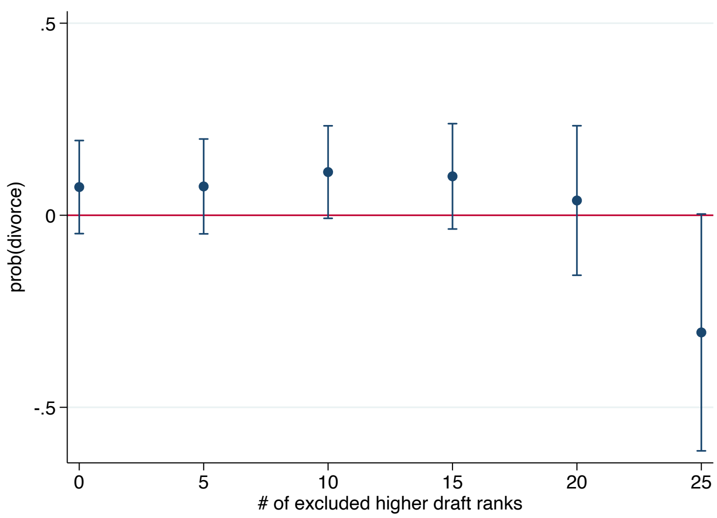


Figure 17: **Propensity to divorce.** This graph illustrates the average causal effect of draft income shock on propensity to divorce for 10 separate regressions: starting with the overall sample (x=0) and progressively restricting to players of lower expected salaries (i.e., excluding the top x=5, 10, 15, 20, 25 draft ranks). Sample restriction ends at x=25 due to the scarcity of divorce (affecting roughly only 8% of ever-married players). Regressions include all controls. Confidence intervals are given at the 95% level. Regressions exclude players who married before the NBA draft.

find.

First, Becker (1973, 1974)'s seminal theory of marriage guides us to think about individuals' wage and time as potential inputs into a family production function $f(x_i, x_j)$ of which the outputs, such as companionship and quantity and quality of children, will determine the economic gains to marriage $\pi_{ij}(f(x_i, x_j))$ for couple ij. Individuals try to maximize their gains to marriage by sorting on traits that improve their joint output. Therefore, we should expect to see positive assortative mating (like marries like) on complementary traits and negative assortative mating (opposites attract) on substitutable traits. Intuitively, wages are substitutes, as a larger couple wage gap increases the comparative advantage of each: The higher earner could specialize more in the labor market and the lower earner could specialize more at home production. The direction in which a change in salary affects a couple's gains to marriage therefore depends on the nature of the couple wage gap. In a case where the man outearns his partner, an increase to the man's wage would widen the couple wage gap, further boost their comparative advantage, and enhance their gains to marriage. On the other hand, an increase to his partner's wage would decrease their wage gap, decrease their comparative advantage, and decrease their gains to marriage.

Keeley (1977) combines Becker's model with standard search theory and predicts that, when the gains to marriage are larger, individuals are more incentivized to enter the marriage market and marry earlier and are more likely to marry overall. Continuing the above example of a couple wage gap where the man earns more than his partner, Becker-Keeley predicts that an increase to the man's wage, in increasing the gains to marriage, would incentivize him to marry earlier, while an increase to the woman's wage, in decreasing the gains to marriage, would cause her to marry later.

This set of Becker-Keeley predictions is precisely what we see in the empirical literature: Women who receive a positive income shock tend to delay marriage, and men who receive a negative income shock tend to be less likely to marry. Conversely, we should expect men who receive a positive income shock to be more likely to marry and marry earlier. My results are therefore consistent with the Becker-Keeley prediction and paint a unifying story for the literature on men.

9 Conclusion

With globally declining marriage rates in recent decades, an important question is whether improvements in male economic status increase marriage rates. However, empirical evidence has proved challenging to attain owing to scant data on positive *per-*

manent income shocks for men. To isolate the causal effect of male earnings on marriage outcomes, I exploit a natural experiment surrounding NBA drafts, where unanticipated changes to player draft order exogenously shift player salary. Constructing a new dataset tracking players' major family decisions, I provide the first empirical evidence that men are indeed more likely to marry when their earnings increase, despite modern-day normalization of cohabitation.

This novel setup yields rich variation in income shocks, revealing a linear causal effect across the domain of positive and negative shocks. Thus, my result not only addresses the gap in our knowledge concerning positive income shocks but also confirms existing results about negative income shocks, providing a unifying story for marriage as a normal good for men. Excluding the superstar draft picks yields effect sizes that are larger and more significant for those with lower expected salaries. Converting these results to income elasticities of marriage demand, my sample's overall elasticity of 0.18 increases to 0.65 after dropping the top 45 draft picks of every cohort. That the elasticity increases for lower expected salaries fits in nicely with Kearney and Wilson (2018)'s general population elasticity of 0.96 from the (1969-1987) Appalachian coal boom. While marriage is a normal good for all three (sub)samples of men, it borders on being a luxury good for general population men who, as expected, are more responsive to changes in income.

I additionally find the income-induced changes in marriage rates are male demand-driven by testing a tabloid theory that NBA players are vulnerable to "baby trapping." I find, however, that players who receive positive draft shocks use their additional bargaining power to effectively *defer* babies until after the engagement or wedding. Not only that, but these income-induced marriages also tend to be high-quality unions not associated with meaningful changes in divorce. Taken together, this is encouraging evidence that, rather than being persuaded into marriage through children, players are choosing to marry in their own time and on their own terms.

Finally, I use the seminal Becker (1973, 1974)-Keeley (1977) framework to tie my results in with those of the literature. That we observe female recipients of positive income shocks tending to delay marriage yet male recipients of positive income shocks tending to embrace it is consistent with their theory of household specialization, given the gender wage gap. That is, changes in earnings that improve the economic gains to marriage induce greater propensity to marry and earlier marriage, and conversely, changes in earnings that erode the gains to marriage reduce likelihood of marriage and delay marriage.

WORKS CITED

Autor, D., D. Dorn, and G. Hanson (2019). When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men. *AER: Insights 1*(2), 161–178.

Baird, S., C. McIntosh, and B. Özler (2011). Cash or Condition? Evidence from a Cash Transfer Experiment. *The Quarterly Journal of Economics 126*(4), 1709–1753.

Becker, G. (1973). A Theory of Marriage: Part I. *Journal of Political Economy 81*(4), 813–46.

Becker, G. (1974). A Theory of Marriage: Part II. *Journal of Political Economy 82*(2), S11–S26.

Black, D. A., N. Kolesnikova, S. G. Sanders, and L. J. Taylor (2013). Are children "normal"? *The Review of Economics and Statistics 95*(1), 21–33.

Business Insider (2011). The Ultimate 2011 NBA Mock Draft Scorecard: Which Hoops Forecaster Did The Best Last Night?

CNBC (2013). The class divide: Marriage as a 'luxury good'.

ESPN (2017). How much the NBA's biggest stars actually earn.

Field, E., R. Glennerster, S. Nazneen, S. Pimkina, and I. Sen (2016). The Effect of Financial Incentives and Girls Empowerment Curriculum on Adolescent Marriage and Schooling in Rural Bangladesh: A Clustered Randomized Controlled Trial. *Duke University Working Paper*.

Forbes (2014). Hooking Up And Marriage: Luxury Goods For Educated Millennials.

Hankins, S. and M. Hoekstra (2011). Lucky in Life, Unlucky in Love? The Effect of Random Income Shocks on Marriage and Divorce. *The Journal of Human Resources 46*(2), 403–426.

Kalmijn, M. (2007). Explaining cross-national differences in marriage, cohabitation, and divorce in Europe, 1990–2000. *Population Studies 61*(3), 243–263.

Kearney, M. S. and R. Wilson (2018). Male Earnings, Marriageable Men, and Nonmarital Fertility: Evidence from the Fracking Boom. *The Review of Economics and Statistics 100*(4), 678–690.

Keeley, M. C. (1977). The Economics of Family Formation. *Economic Inquiry 15*(2), 238–250.

Kennedy, Alex (host) (2017). Darington Hobson gives inside look at each step of NBA draft process. *HoopsHype Podcast*.

Lee, G. R. and K. K. Payne (2010). Changing Marriage Patterns Since 1970: What's Going On, and Why? *Journal of Comparative Family Studies 41*(4), 537–555.

Pew Research Center (2017). 2017 American Trends Panel.

Pew Research Center (2019). Marriage and Cohabitation in the U.S.

Salisbury, L. (2017). Women's Income and Marriage Markets in the United States: Evidence from the Civil War Pension. *The Journal of Economic History 77*(1), 1 – 38.

Schoen, R. and V. Canudas-Romo (2005). Timing effects on first marriage: Twentieth-century experience in England and Wales and the USA. *Population Studies 59*(1), 135–146.

Schoen, R. and A. Cheng (2006). Partner Choice and the Differential Retreat From Marriage. *Journal of Marriage and Family 68*(1), 1–10.

Teitler, J. O., N. E. Reichman, L. Nepomnyaschy, and I. Garfinkeld (2009). Effects of Welfare Participation on Marriage. *Journal of Marriage and Family 71*(4), 878–891.

The Wall Street Journal (2020). Affluent Americans Still Say 'I Do.' More in the Middle Class Don't.

U.S. Census Bureau (2020). Estimated Median Age at First Marriage, by Sex: 1890 to the Present.

A Salary Data: International estimates

Team payroll budgets are collected from an assortment of sources, from archived press releases to annual sports magazine publications of team budgets and budget rankings lists, cross-checked with interbasket.net forum contributions by avid budget trackers as crowd-sourced due diligence for disclosure accuracy. This cross-checking is important as it is well known that some teams may announce an inflated budget to sound more impressive, and conversely, some sports magazines may understate the team payrolls of their nation to boost perceived return on investment relative to other nations' basketball teams. Additionally, forum discussion helps clarify any ambiguity regarding whether announced 'budgets' are 'total' or 'payroll' budgets based on the affordability of the team's roster of players. Payroll is usually 50% of the total budget, but this percentage tends to vary by nation. For example, France is on the low end with payroll budget at around 33% of total budget, and Spain is on the higher end at roughly 75%. In instances where only a team's total budget is known, I approximate the team's payroll budget using the team's payroll percentage. If the team's payroll percentage is also unknown, I use the average of the team's nation's known payroll shares.

Next, I use proballers.com to obtain team rosters and each player's average minutes per game (MIN stat). I then use player MIN stats to calculate each player's payroll share (as player MIN divided by aggregate MIN), given that MIN is greater than zero for all players and larger for more valuable players, accounts well for player salary variation in the payroll. This payroll share represents each player's potential (full season) contract value. While actual contract value is generally the full contract value, I calculate partial contracts for players who are on monthly contracts or leave early due to mutual release agreements, being fired, or team bankruptcy by comparing player employment duration with the league's season duration. If budget information cannot be found, generally for relatively unknown teams, I collect typical overseas salaries by continent and team tier.

B Family Data: Sources

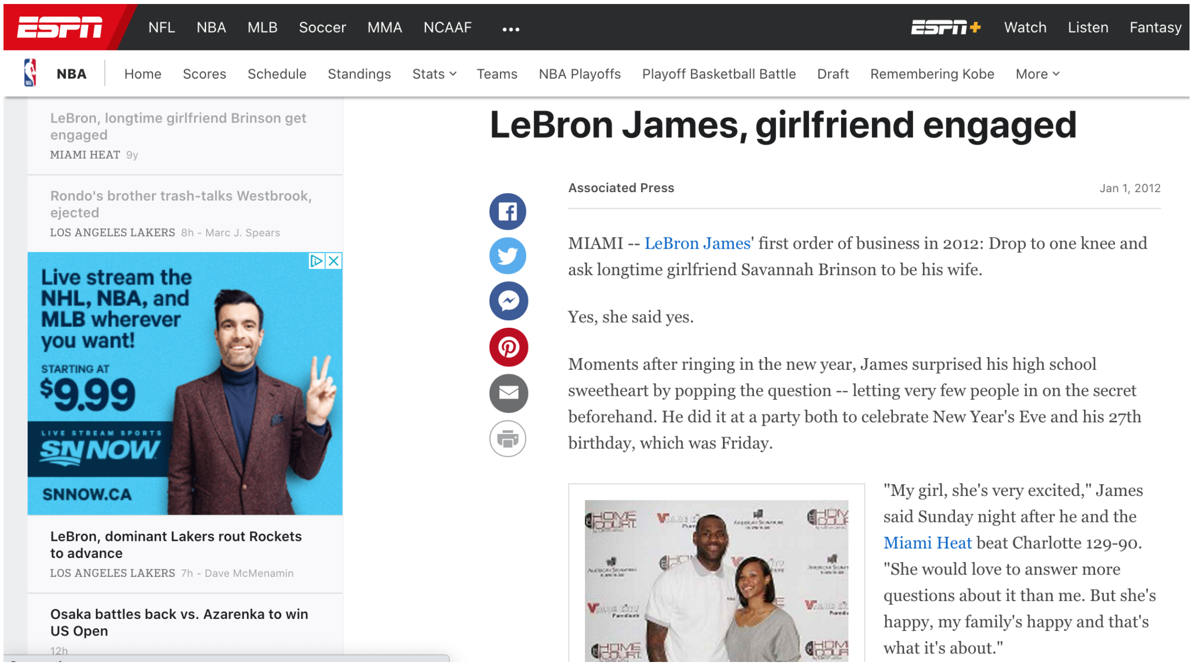
To give an idea of the data collection process, this section provides examples of how family data typically appears in the lens of social media and sports interviews, explores potential pitfalls to automation strategies that we learn along the way, and includes a non-English sample for overseas players.

B.1 Example: social media and sports articles

For familiarity and verifiability, let us examine for this exercise the family of NBA household name LeBron James. The following Instagram post, for example, tells us that LeBron has an Instagram-verified (blue check-marked) account (kingjames) and that 2018 marks not only 5 years of marriage with his wife Savannah but also a relationship milestone of 18 years. Given this, we can back out their marriage year to be 2013 and that their relationship began in 2000 before his 2003 NBA draft. Savannah's account (mrs_savannahrj) bio further tells us that she is a "Wife · Mother · Daughter · Sister · Friend · Philanthropist · Business Woman."

Their engagement year is provided in the following 2012 ESPN article. Despite the proposal taking place during a 2011 New Year's Eve party, the article clarifies that LeBron in fact proposed just after midnight on Jan 1, 2012. The article further confirms that the couple started dating in high school, before LeBron was drafted into the NBA.

Furthermore, in absence of common knowledge on child birthdays or articles listing the ages of all the children, birth years can be deduced on a child-by-child basis through birthday shoutouts. The birth year of the James family's youngest child, Zhuri Nova Ann-Marie (allthingszhuri), for example, can be deduced to be 2014 from the following 2019 Instagram post for her 5th birthday. The birth years of his two sons can be found



LeBron James, girlfriend engaged

Associated Press Jan 1, 2012

MIAMI -- LeBron James' first order of business in 2012: Drop to one knee and ask longtime girlfriend Savannah Brinson to be his wife.

Yes, she said yes.

Moments after ringing in the new year, James surprised his high school sweetheart by popping the question -- letting very few people in on the secret beforehand. He did it at a party both to celebrate New Year's Eve and his 27th birthday, which was Friday.

"My girl, she's very excited," James said Sunday night after he and the Miami Heat beat Charlotte 129-90. "She would love to answer more questions about it than me. But she's happy, my family's happy and that's what it's about."

in a similar fashion.

While most players have public social media accounts that are visible to their fans, some ingenuity is required to find some accounts. Besides the obvious that some players may register under a nickname or alternative spelling of their legal name, navigation around impostor accounts is oftentimes necessary, such as checking to see if the account is followed by fellow team mates. Some players may be elusive and omit their name entirely, rendering their account not directly searchable.[32] In these cases, we can source the knowledge of the crowds by running a hashtag search of the player's full name. This often returns action footage and photos posted by basketball teams, media outlets, and avid fans, who can often be counted on to know the player's username and tag his account. Once found, note that besides checking the photos posted by the player himself, it is just as important to check the photos in which he is tagged by others—especially if the player does not tend to post family photos himself—as oftentimes, his partner may be the one actively tagging him in photos of their children. Finally, while there are some players with private accounts, the vast majority of them accept 'follow' requests, which helps to reduce missing observations.

[32] In the above example, LeBron's account (kingjames) currently has his full name 'LeBron James' but no bio, and Savannah's account (mrs_savannahrj) has a bio but no name, meaning that her account would not appear in a search for 'Savannah Brinson.' Their daughter Zhuri's account (allthingszhuri) has both a name and a bio, except the name used is "All Things Zhuri."


kingjames • Following
kingjames Happy 5th Beautiful-Day party(Tues official day/date) my Princess 👸 Zhuri Nova Ann-Marie James!!!!!! Daddy loves his lil girl with every single inch and more of his heart for eternity and beyond! Enjoy your day mommy face❤️❤️❤️❤️❤️❤️❤️❤️❤️❤️❤️❤️ #JamesGang👑
47w
postal86 She looks just like your beautiful wife❤❤❤
30w 3 likes Reply
Liked by slim_niv34 and 935,634 others
OCTOBER 19, 2019
Add a comment...
Post

B.2 Potential pitfalls of automation

In the example above with LeBron James, we can already note several potential pitfalls to algorithmic automation of data scraping in each of the sample sources.

First, in LeBron's anniversary post, key words such as 'wedding' or 'marriage' are not explicitly stated and must be deduced from context. Furthermore, the five years of marriage he is celebrating with his wife is denoted not by text but by a hand emoji, and the only mention of 'years' in his post is referring to a different (dating) anniversary. Not only is the use of emojis – especially as a substitute for key words – prevalent in social media posts, many carry multiple meanings (as seen with 'hand' versus 'five'), which can be difficult for an algorithm to parse, leading to missed data collection opportunities. Also, collecting key numbers at a cursory glance (such as "18 years" in a wedding anniversary post) without checking for context could falsely attribute the year they started dating (2000) to the year in which they got married (2013).

Next, in the ESPN engagement article, simply scanning for key facts would yield conflicting results: while 'New Year's Eve' implies December 2011, the article time stamp is January 2012. In this case, the article year of 2012 gives us the correct engagement year but only because LeBron happened to propose "moments after ringing in the new year." Therefore, simply collecting the publication year of the article without scanning its contents is error prone due to lag time in writing and publication. Furthermore, searching for keywords in the articles has its own pitfalls. Feature articles on weddings, for example, often mention other team mates or family in attendance. Therefore, it is not enough that a player's name appears in a marriage article but that there needs to be a check to ensure the player in question is the groom and not simply a wedding guest at someone else's wedding.

Finally, in LeBron's post for his daughter Zhuri's birthday, the keyword 'birthday' is never explicitly stated as LeBron instead refers to her special day as her "Beautiful Day," the significance of which, again, must be deduced from context. Others may even leave their photo captions blank, in which case the photo itself needs to be scanned for clues. Alternatively, key words could be found in a friend/family's comment or the player's answer to a someone's question. However, simple keyword search is again error prone as context is required to differentiate between comments and questions. For example, comments like "I can't believe little Rupert is already 1!" helps match birth year to child name. On the other hand, a fan asking "Is that your son?" on an uncaptioned photo of a player spending time with his nephew should not cause the photo to be erroneously flagged as evidence of having biological children just because the keyword 'son' is present.

B.3 Example: foreign articles

The data collection process for international players is similar except involving a plethora of foreign languages. Loukas Mavrokefalidis (57^{th} pick, 2006 draft), for example, has no English search results about his family life. Since Mavrokefalidis is not only ethnically Greek but also played mostly for Greek basketball teams, such as Olympiacos and Marousi B.C., it makes sense to first search for Greek articles using the same keyword search methodology, except in Greek. First, I look up Loukas' unromanized Greek name (Λουκάς Μαυροκεφαλίδης), for which Wikipedia is a great go-to source. Searching his name in combination with keywords such as 'wedding' (γάμος) or 'wife' (γυναίκα), provided by Google Translate, yields the following Greek Gazette article, announcing Loukas's 2017 wedding ceremony with Angeliki (Αγγελική) but also clarifying that they already had their civil marriage in 2011. In this case, I record the marriage year to be the earlier of the two, 2011. Child birth year data can be found similarly by including his wife's name in the search or keywords such as 'son' (υιός), 'daughter' (κόρη), and 'children' (παιδιά).

Παντρεύτηκε την εκλεκτή της καρδιάς του με παπά και με κουμπάρο ο Μαυροκεφαλίδης

Την αγαπημένη του Αγγελική παντρεύτηκε με θρησκευτικό γάμο ο Λουκάς Μαυροκεφαλίδης στο Κιλκίς.

SHARE

CONTRA.GR

04 ΙΟΥΛ 2017

Μπορεί να ήταν παντρεμένος με πολιτικό γάμο εδώ και μερικά χρόνια, όμως ήρθε η ώρα της... εκκλησίας.

Apart from articles and the player's own social media, another great source of data is the partner's social media. For example, while Nando de Colo (58$^{\text{th}}$ pick, 2009 draft) provides ample posts about his wedding day and periodic family updates, he did not post about his engagement. His wife, Vero Compañ, however, did post a public photo commemorating their engagement. While her photo is uncaptioned and features both a baby bump and an engagement ring, we can deduce from her friends' comments about her engagement ring (anillo de compromiso) and earlier photos' pregnancy announcement that this photo indeed constitutes an engagement announcement. From context, this is not a throwback photo, and thus the time stamp tells us their engagement year was 2014.